A social and neurobiological approach to pointing in speech and gesture

David Peeters

ISBN 978-90-76203-69-0

Printed and bound by Ipskamp Drukkers, Nijmegen.

The research reported in this dissertation was supported by an International Max Planck Research School fellowship funded by the Max Planck Society, Munich, Germany.

A social and neurobiological approach to pointing in speech and gesture

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. Th.L.M. Engelen,
volgens besluit van het college van decanen
in het openbaar te verdedigen op maandag 14 september 2015
om 14.30 uur precies

door
David Gerard Theodoor Peeters
geboren op 25 april 1987
te Eindhoven

Promotoren:

Prof. dr. A. Özyürek

Prof. dr. P. Hagoort

Manuscriptcommissie:

Prof. dr. T. Dijkstra

Prof. dr. S. Kita (University of Warwick, Groot-Brittanië)

Dr. T. C. Gunter (Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Duitsland)

A social and neurobiological approach to pointing in speech and gesture

Doctoral Thesis

to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. Th.L.M. Engelen,
according to the decision of the Council of Deans
to be defended in public on Monday, September 14, 2015
at 14.30 hours

by
David Gerard Theodoor Peeters
Born on April 25, 1987
in Eindhoven, The Netherlands

Supervisors:

Prof. dr. A. Özyürek
Prof. dr. P. Hagoort

Doctoral Thesis Committee:

Prof. dr. T. Dijkstra
Prof. dr. S. Kita (University of Warwick, United Kingdom)
Dr. T. C. Gunter (Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany)

Table of contents

Chapter 1

General Introduction

General Introduction

> "Deixis introduces subjective, attentional, intentional and of course context-dependent properties into natural languages"
> (Levinson, 2004, p. 97)

We refer to the things around us. In everyday communication, we often use words to describe intended referents, and our bodies (e.g., eyes, head, hands, and torso) to indicate the location to which our addressee should focus her attention in order to further identify what we are talking about (Bühler, 1934; Clark & Bangerter, 2004). Prima facie this is a very simple and straightforward part of human communication. Upon careful inspection however, simple acts of reference turn out to require quite a complex interplay between speaker and addressee, arguably relying on multiple cognitive mechanisms and multimodal social cues. In a prototypical instance of successful everyday referential communication, a speaker produces a manual pointing gesture to a physical object, often in temporal alignment with a spoken referential expression that canonically contains a spatial demonstrative (as in *I have bought <u>that</u> book*), while alternating gaze between addressee and referent. At the same time, the addressee perceives the speech, gesture, and other bodily behavior of the speaker, integrates the transmitted visual and auditory information, recognizes and understands the speaker's communicative intention and social motive, and shifts her gaze to identify the referent and establish joint attention.

In instances of demonstrative reference, speaker and addressee thus need to work together in a collaborative and interactive process to establish a joint focus of attention on a referent (Clark & Bangerter, 2004; Clark & Wilkes-Gibbs, 1986; Tomasello, Carpenter, & Liszkowski, 2007). It can therefore be considered a joint action (Clark, 1996) and a clear instance of human cooperative communication (Tomasello, 2008).

The current thesis focuses on such triadic situations in which speech and gesture are used in a *deictic* manner, i.e. they relate to the space (and time) in which they occur and their interpretation is for

an important part dependent on the context in which they are produced (Bühler, 1934; Levinson, 1994; McNeill, 1992). Deixis (or: 'indexicality') has been a topic of interest in many different scientific disciplines, including philosophy (e.g., Peirce, 1955; Russell, 1940; Quine, 1960; Wittgenstein, 1953), linguistics and anthropology (Enfield, 2003; Hanks, 1990; Levinson, 1983), and developmental and comparative psychology (e.g., Clark & Sengul, 1978; Henderson, Yoder, Yale, & McDuffie, 2002; Küntay & Özyürek, 2006; Leavens, Hopkins, & Bard, 1996; Tomasello et al., 2007). However, in spite of the complexity of everyday demonstrative reference outlined above, linguistic (e.g. spatial demonstratives) and gestural (e.g. pointing gestures) markers of demonstrative reference have often been studied in isolation. Furthermore, the neurocognitive mechanisms that allow us to refer to the things around us and those that allow us to understand our interlocutors' referential acts remain largely unclear.

The current thesis therefore brings spatial demonstratives and pointing gestures together in an experimental and neurobiological approach to demonstrative reference. It studies the phenomenon from both the production and the comprehension side and takes into account the social and communicative context in which demonstrative reference generally occurs. This multifaceted approach allows contrasting and testing different theoretical views of spatial deixis (see below) and advances our understanding of the social, cognitive and neural mechanisms underlying demonstrative reference. In general, "any analytical inquiry is destined to compartmentalize different parts of a complex system" (Kita, 2003, p. 307). The different chapters of this thesis do exactly that by focusing on different elements of the larger complex system of everyday demonstrative reference in a visual, triadic, and communicative context.

In the remainder of this General Introduction I will first outline the traditional and dominant view of spatial deixis, in which demonstratives have played a central role. Second, I will discuss an alternative, sociocentric view. Third, I will elaborate on the approach taken in the current thesis to contrast these two views in four empirical studies. Finally, I will provide an overview of the structure of

the thesis and briefly clarify the specific experimental methods that are used in each of the empirical chapters.

The null hypothesis: Egocentric proximity

Traditional views of spatial deixis, across scientific disciplines and generally inspired by Bühler's (1934) theory of language, are *egocentric* in nature. They often focus on spatial demonstratives such as *this* and *that* and argue that by using these terms speakers "indicate the relative distance of an object, location, or person vis-à-vis the deictic center (also called the *origo*), which is usually associated with the location of the speaker" (Diessel, 1999, p. 36). In other words, "the anchoring point of deictic expressions is egocentric (or, better, speaker-centric). Adult speakers skillfully relate what they are talking about to this me-here-now" (Levelt, 1989, p. 46), such that "the speaker, by virtue of being the speaker, casts himself in the role of ego and relates everything to his viewpoint" (Lyons, 1977, p. 638). In the case of simple two-term demonstrative systems, this means that, canonically, a *proximal* demonstrative is used for referents relatively close to the speaker, and a *distal* demonstrative for referents relatively remote from where the speaker is located. This egocentricity in how demonstratives are used is said to be a universal property of demonstrative systems. Diessel (2014, p. 128), for instance, states that "speakers of all languages employ an egocentric coordinate system that is anchored by the speaker's body at the time of the utterance" (see also Levelt, 1989, p. 47). The egocentric proximity-based view has been very influential, is intuitively appealing, and is still present in the literature today (e.g., Anderson & Keenan, 1985; Clark & Sengul, 1978; Coventry, Valdés, Castillo, & Guijarro-Fuentes, 2008; Diessel, 2005, 2014; Fillmore, 1982; Halliday & Hasan, 1977; Hottenroth, 1982; Lakoff, 1974; Levelt, 1989; Lyons, 1977; Rauh, 1983; Russell, 1940; Stevens & Zhang, 2013). Surprisingly, very few such studies take into account the interplay between spatial demonstratives and pointing gestures in acts of multimodal reference.

In addition, this theoretical view of deixis focuses on the speaker so much that it often does not consider how addressees comprehend or interpret the demonstratives they hear. However, according to Diessel (2014), demonstratives are not only used but also interpreted (by an addressee) based on the relative distance to the deictic center (i.e. the speaker) in face-to-face conversations. An addressee will thus generally expect that a speaker uses a proximal demonstrative in reference to an object, location, or person that is relatively close to the speaker's body at the time of the utterance and a distal term for entities that are relatively further away from the speaker (cf. Stevens & Zhang, 2013). The straightforward predictions made by this account, with regards to both the production and comprehension of demonstrative reference, will serve as the null hypothesis that is tested in this thesis. The alternative hypothesis, described below, considers demonstrative reference a sociocentric, collaborative, and multimodal phenomenon.

The alternative hypothesis: Sociocentric collaboration

Traditional accounts of demonstrative reference argue that speakers are egocentric and that they mainly use demonstrative terms as a function of the physical proximity of the referent. The alternative is that demonstrative reference is a sociocentric phenomenon (Hanks, 1992) in which the addressee plays a crucial role and in which contextual factors other than the relative proximity of the referent drive the demonstrative choice. In-depth observational studies of demonstrative use in everyday interactions in different languages suggest that this may be the case. Enfield (2003), for instance, in describing the Lao two-term demonstrative system, concludes that "distance cannot be what distinguishes the meanings of these two demonstratives" (p. 104). Rather, demonstrative reference is described as a social, interactive process in which the choice for a proximal or distal demonstrative depends on how interlocutors perceive and interpret the physical space during their interaction (Enfield, 2003). What is perceived as

"proximal" may depend, for instance, on the engagement areas of speaker and addressee during their conversation (Enfield, 2003).

Other work also suggests that demonstrative reference is a sociocentric phenomenon in which social factors are more important than the relative distance of a referent in a speaker's demonstrative choice. Piwek, Beun, and Cremers (2008), for instance, argue that demonstrative choice in Dutch is not driven by the relative proximity of a referent to the speaker, but by the cognitive and visual accessibility of a referent to the speaker and the addressee. Özyürek (1998) argues that the Turkish demonstrative *şu* is used as a function of whether the addressee already looks at the intended referent or not. Jungbluth (2003) states that the physical orientation of interlocutors relative to each other in a conversation drives demonstrative choice in Spanish. When speaker and addressee are face-to-face in a conversational dyad, all referents within the dyad are treated as proximal "without any further differentiation" (Jungbluth, 2003, p. 19). Thus, possible referents are not considered from an egocentric point of view, but from a sociocentric perspective. More than traditional views of spatial deixis, such sociocentric accounts are in line with the broader perspective of everyday reference as a collaborative, joint action in which speaker and addressee work together to establish a joint focus of attention. The current thesis contrasts these two views of spatial deixis in both production and comprehension.

The current thesis

An experimental and neurobiological approach to spatial deixis

The theoretical differences between egocentric and sociocentric accounts of spatial deixis may be related to the different types of methodological approach that have been taken in the study of spatial demonstratives. The egocentric view is largely based on linguistic intuitions (or: the 'armchair' approach, Clark & Bangerter, 2004), which have often shaped descriptions of demonstrative systems in typological sources such as reference grammars (see Diessel, 2005, for an overview). Indeed, when

asking a naïve informant how s/he uses the terms “this” and “that” in daily conversations, the reply will almost always be related to the physical distance of referents to the speaker’s physical location. Moreover, when participants are explicitly asked to judge whether a speaker uses a demonstrative correctly, they will take the distance of a referent to the speaker as the criterion on which they base their judgments (Stevens & Zhang, 2013). One may question, however, whether such linguistic intuitions are reliable. Extensive analysis of demonstrative choice in different languages on the basis of (videotaped) observational materials has shown that people’s intuitions are often simply not in line with their actual patterns of demonstrative use in everyday interactions (e.g., Enfield, 2003; Hanks, 1990; Özyürek, 1998).

More recently, experimental approaches have aimed to shed more light on demonstrative choice, for instance by placing objects at different distances from a participant in the lab and eliciting the production of demonstratives (e.g., Coventry et al., 2008; Coventry, Griffiths, & Hamilton, 2014; Stevens & Zhang, 2014). An important advantage of experimentally studying demonstrative reference is that insights from observational work can be tested and further specified in a controlled, experimental environment, which allows disentangling the relative influence of different variables (e.g. attentional, cultural, intentional, physical, social, subjective) that are not easily distinguishable in everyday multi-modal communication. The current thesis takes such an experimental approach to contrast and test egocentric and sociocentric theories of spatial deixis, aiming to advance our understanding of the production and comprehension of spatial demonstratives. In doing so, it builds on knowledge obtained from earlier observational studies and goes beyond previous experimental work in studying the phenomenon from a multimodal perspective and from both the production and the comprehension side. As we will see, many experimental findings reported in this thesis align well with and further specify suggestions that were made by observational work. Importantly, however, they also clearly reject other conclusions that were based on linguistic intuitions.

Furthermore, the current thesis makes use of methods from cognitive neuroscience to study demonstrative reference. This novel, neurobiological approach has at least two advantages. First, it allows putting different theoretical views to the test by inspecting the electrophysiological signatures of the production and comprehension of demonstrative reference in different experimental conditions. Second, the neural and cognitive underpinnings of demonstrative reference themselves, in both production and comprehension, have remained largely unclear. Therefore, a neuroscientific approach to the phenomenon advances our understanding of the neurobiological mechanisms involved in everyday referential acts, for instance by identifying the brain regions involved in the perception and integration of auditory and visual markers of multimodal reference to an object in a triadic context.

In addition to taking an experimental and neurobiological approach to the study of demonstrative reference, the current thesis also acknowledges the multimodal nature of the phenomenon. As outlined in the following section, it thus studies the use and comprehension of demonstratives not in isolation, but in a broader, visual context in which speakers of different languages use their body to indicate which referent they want their addressee to focus on, for instance by producing a pointing gesture.

A multimodal approach to spatial deixis

A core property of human communication is that it allows us to shift the attention of our conversational partners to entities in the world around us, and often, unlike our closest phylogenetic relatives, we do so simply to share interest in a particular referent with one another (Clark, 1996; Kita, 2003; Tomasello et al., 2007). As we have seen above, spatial demonstratives often form an important part of the spoken component of a multimodal utterance that is produced to establish a joint focus of attention. Such linguistically transmitted information is often paired with some form of pointing, for instance with the head, chin, thumb, or the extended index-finger (see Cooperrider & Núñez, 2012; Enfield, 2001; Kendon, 2004; Kita, 2003; Sherzer, 1973; Wilkins, 2003). Although it is generally

acknowledged that both demonstratives and pointing gestures play a pivotal role in shifting the addressee's attention to a referent (e.g., Diessel, 2006), surprisingly, they have often been studied as separate phenomena (but see Bangerter, 2004; Cooperrider, 2011). Links that have been made remain largely descriptive. For instance, it has been found that some demonstratives need an accompanying pointing gesture (e.g., Senft, 2004), whereas others are rarely accompanied by pointing (e.g., Burenhult, 2003). Moreover, within languages some demonstratives are paired with a pointing gesture more commonly than others (Cooperrider, 2011; Küntay & Özyürek, 2002; Piwek et al., 2008). The current thesis acknowledges that demonstrative reference is a multimodal phenomenon (cf. Bangerter, 2004) and aims to use insights about pointing to inform theories of spatial deixis that have focused mainly on demonstratives.

In producing a spatial demonstrative, speakers of most languages have the choice between different linguistic alternatives (Diessel, 2005). Similarly, the exact form a pointing gesture takes is variable and not fully determined a priori (De Ruiter, 2000). Different articulators may be used, and parameters such as the trajectory, endpoint, and velocity of the gesture may differ across instances. If demonstrative reference is considered an egocentric act, it is unlikely that people will modify their gesture on the basis of social considerations that relate to their addressee. In other words, irrespective of the speaker's communicative intentions, the gesture will be the same (cf. Brunetti et al., 2014). Under such an account, it has been considered sufficient to study the kinematic properties and neurobiological underpinnings of pointing gestures in the lab in situations in which there is no addressee involved (see Cleret de Langavant et al., 2011, for discussion).

Alternatively, people may design the form of their pointing gestures from a sociocentric perspective, and there are some preliminary indications that this is the case. Enfield, Kita, and De Ruiter (2007), for instance, distinguish between relatively big points in which the whole arm is outstretched and relatively small points in which the hand is the main articulator. Big points would do the primary

work of an utterance, such as pointing out the location of an object, whereas small points would occur in utterances in which speech is central, adding a background modifier on the basis of social and communicative factors such as the common ground between interlocutors (p. 1738). Cleret de Langavant et al. (2011) report pointing gestures that had a trajectory and endpoint distribution that were tilted away from the gesturer's addressee, arguably because the addressee's perspective on the target object was taken into account in the form of the gesture.

One of the aims of the current thesis, therefore, is to examine whether and how the form (or: kinematics) of index-finger pointing gestures differs as a function of one's communicative intentions and as a function of the presence of concomitantly produced spatial demonstratives. Models of speech production (Levelt, 1989) and speech and gesture production (e.g., De Ruiter, 1998; Kita & Özyürek, 2003) stress the importance of communicative intentions in driving the production of (multimodal) utterances. However, the exact influence of intentions on this type of action remains unclear. If speakers are egocentric in demonstrative reference, no difference in the kinematics of their pointing gesture is expected as a function of their communicative intentions. The sociocentric alternative is that speakers tailor the kinematics of their gesture to the needs of their addressee.

Outline and methodology

This thesis presents four empirical studies that focus on different aspects of the multimodal act of establishing joint attention to a referent in a visual, triadic context. Together these studies contrast and test egocentric and sociocentric views of spatial deixis and aim to further our understanding of the cognitive and neural underpinnings that support the production and comprehension of demonstrative reference. Chapters 2 and 3 focus on spatial demonstratives in a larger multimodal context including pointing gestures. Chapters 4 and 5 investigate pointing gestures in a wider multimodal context including demonstrative speech.

Chapter 2 focuses on the *production* of spatial demonstratives in Dutch and Turkish. Observational work in both languages suggests that contextual factors beyond the physical proximity of a referent influence speakers' demonstrative choice. Building on such observational findings, this chapter further investigates in a controlled setting whether and to what extent the attentional status of the addressee, the location of a physical referent, and the presence of a manual pointing gesture influence the particular demonstrative native speakers of Dutch and Turkish use in triadic settings. A controlled elicitation task was developed that, going beyond the methods generally used in observational work, allowed orthogonal manipulation of these different contextual factors and observation of their individual and interacting contributions to demonstrative choice in both languages. The egocentric account described above predicts that only the relative proximity of a referent to the speaker determines demonstrative choice. An influence of other contextual factors such as the locus of attention of the addressee is in line with a sociocentric view of demonstrative reference.

Chapter 3, on the other hand, focuses on the *comprehension* of spatial demonstrative terms. Two experiments are presented in which participants' electrophysiological brain activity was recorded while they saw pictures of a person pointing at an object and listened to her referential speech. These experiments allowed for directly contrasting and testing, from a comprehension perspective, the two main theoretical views on demonstratives that were described above. In addition, they further our understanding of how the brain allows one to comprehend and integrate demonstrative speech and gesture in a visual, everyday context. Studies investigating the comprehension of spatial demonstratives are scarce, and an important novelty of this chapter is that it takes into account how the physical, spatial orientation of speaker and addressee vis-à-vis each other may influence the addressee's demonstrative comprehension in relation to the location of referents. The egocentric proximity account predicts that participant addressees interpret spatial demonstratives as a function of the relative distance of a referent to the speaker (e.g., Diessel, 2014). The sociocentric alternative is that socially relevant

factors such as the position of a referent inside or outside the shared space between speaker and addressee play a more important role (e.g., Jungbluth, 2003), as reflected in participants' electrophysiological brain response time-locked to hearing a demonstrative.

In Chapter 3 participants' electrical brain activity is measured by electrodes placed in a cap covering the scalp, a method known as electroencephalography (EEG). The electrophysiological correlates of brain activity recorded in this way are time-locked to important events in the experiment (such as the onset of critical stimuli), baseline-corrected, filtered, and averaged off-line, which creates event-related potentials (ERPs). In Chapter 3, the assumption is that differences in processing ease of different demonstrative terms in a particular visual context should be reflected in the ERPs. The logic behind this approach is inspired by "regular" N400 experiments in which the contextually incongruent use of a noun elicits a negative deflection in the N400 component relative to its congruent counterpart (see e.g., Kutas & Federmeier, 2011; Kutas & Hillyard, 1980). However, in regular N400 experiments, congruent and incongruent experimental conditions are generally defined before the start of the experiment on the basis of the intuitions of the researcher or the results of a pre-test. In the case of spatial demonstratives, there are different theoretical views on what it means that a demonstrative is used in a 'congruent' or 'incongruent' way, and meta-linguistic intuitions seem unreliable in this respect. Therefore, the simple recording of the brain's electrophysiological response to hearing different demonstrative terms in a visual, triadic context allows the contrasting of opposing theoretical views to shed more light on which demonstrative term an addressee prefers and expects in a particular context.

Chapter 4 focuses on the planning and *production* of index-finger pointing gestures in the context of demonstrative reference. More specifically it investigates the role of one's communicative intentions in shaping unimodal and bimodal acts of pointing as a function of the shared knowledge between speaker and addressee. Two experiments are presented in which participants produced index-finger pointing gestures in the lab, while their index-finger kinematics and electrophysiological brain

activity were continuously recorded. This approach allows for the exploration of how people may use the kinematics of their pointing gesture in order to be more or less informative as a function of their communicative intentions. In addition, it explores whether and when intentional and/or attentional neurocognitive mechanisms are involved in the planning of a communicative pointing gesture. The egocentric null hypothesis is that speakers do not consider the addressee's knowledge state while shaping the kinematics of their gesture. The sociocentric alternative is that they modulate the kinematics of their gesture to be as informative as necessary for their addressee and that this is reflected in the neuronal activity preceding the execution of their gesture.

The recording of index-finger kinematics in Chapter 4 is allowed by the use of a motion tracking system that continuously records the spatial position of a marker placed on the nail of participants' index-finger. By comparing the spatial location of the marker at different time points during the execution of pointing gestures, the velocity and duration of different components of the gesture (e.g., the stroke and post-stroke hold phase) could be calculated. At the same time, participants' EEG is continuously recorded, which allows for pinpointing electrophysiological markers of different cognitive mechanisms involved in the planning and production of gesture.

Chapter 5 presents a functional magnetic resonance imaging (fMRI) study investigating the neural integration of index-finger pointing gestures and referential speech in *comprehension*. The majority of previous studies that investigated the neural correlates of speech-gesture integration looked at *iconic* gestures. However the neural infrastructure involved in integrating speech and pointing gestures is currently unclear. Therefore, participants were placed inside an MR scanner and watched images of a speaker pointing at an object while listening to her speech that referred to the object she pointed at. The blood oxygenation level dependent (BOLD) response to bimodal (speech + pointing gesture) presentation of stimuli was compared to the sum of unimodal (speech-only and pointing gesture-only) presentations of the same stimuli. In addition, a matching condition (i.e. the speaker

correctly named the object she pointed at) was compared to a mismatch condition (i.e. the speaker named one object but pointed at another). Both these manipulations allow for investigating the putative role of bilateral superior/middle temporal and left inferior frontal regions of the brain in the semantic unification (Hagoort, 2005; 2013) and audiovisual integration of speech and pointing gestures in comprehension.

Finally, **Chapter 6** provides a summary of the four empirical chapters and discusses the findings from the broader perspective of multimodal demonstrative reference.

References

Anderson, S. R., & Keenan, E. L., (1985). Deixis. In T. Shopen (Ed.), *Language typology and syntactic description* (pp. 259-308). Cambridge: Cambridge University Press.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*(6), 415-419.

Brunetti, M., Zappasodi, F., Marzetti, L., Perrucci, M. G., Cirillo, S., Romani, G. L., Pizzella, V., & Aureli, T. (2014). Do you know what I mean? Brain oscillations and the understanding of communicative intentions. *Frontiers in Human Neuroscience*, *8*, 36.

Bühler, K. (1934). *Sprachtheorie*. Jena: Fischer.

Burenhult, N. (2003). Attention, accessibility, and the addressee: The case of the Jahai demonstrative ton. *Pragmatics, 13*, 363-379.

Clark, E. V., & Sengul, C. J. (1978). Strategies in the acquisition of deixis. *Journal of Child Language*, *5*(3), 457-475.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Bangerter, A. (2004). Changing ideas about reference. In I. A. Noveck, & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 25-49). Basingstoke: Palgrave Macmillan.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39.

Cleret de Langavant, L., Remy, P., Trinkler, I., McIntyre, J., Dupoux, E., Berthoz, A., & Bachoud-Lévi, A. C. (2011). Behavioral and neural correlates of communication via pointing. *PloS one*, *6*(3), e17719.

Cooperrider, K. (2011). Reference in action: Links between pointing and language. Doctoral dissertation, University of California, San Diego.

Cooperrider, K., & Núñez, R. (2012). Nose-pointing: Notes on a facial gesture of Papua New

Guinea. *Gesture*, *12*(2), 103-129.

Coventry, K. R., Griffiths, D., & Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: Describing and remembering object location. *Cognitive psychology*, *69*, 46-70.

Coventry, K. R., Valdés, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language within your reach: Near–far perceptual space and spatial demonstratives. *Cognition*, *108*, 889-895.

De Ruiter, J. P. (1998). *Gesture and speech production.* Doctoral dissertation, University of Nijmegen, The Netherlands.

De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture* (pp. 284-311). Cambridge: Cambridge University Press.

Diessel, H. (1999). *Demonstratives. Form, Function, and Grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2005). Distance contrasts in demonstratives. In M. Haspelmat, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The World Atlas of Language Structures* (pp. 170-173). Oxford: Oxford University Press.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, *17*(4), 463–489.

Diessel, H. (2014). Demonstratives, Frames of Reference, and Semantic Universals of Space. *Language and Linguistics Compass*, *8*(3), 116-132.

Enfield, N. J. (2001). 'Lip-pointing': A discussion of form and function with reference to data from Laos. *Gesture*, *1*(2), 185-211.

Enfield, N. J. (2003). Demonstratives in space and interaction: Data from Lao speakers and implications for semantic analysis. *Language*, 82-117.

Enfield, N. J., Kita, S., & De Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, *39*(10), 1722-1741.

Fillmore, C. J. (1982). Towards a descriptive framework for spatial deixis. In R. J. Jarvella, & W. Klein (Eds.), *Speech, place, & action. Studies in deixis and related topics* (pp. 31-59). Chichester: John Wiley & Sons Ltd.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in cognitive sciences*, *9*(9), 416-423.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in psychology*, *4*.

Halliday, M. A. K., & Hasan, R. (1977). *Cohesion in English*. London, UK: Longman Group Ltd.

Hanks, W. F. (1990). *Referential practice: Language and lived space among the Maya*. Chicago: University of Chicago Press.

Hanks, W. F. (1992). The indexical ground of deictic reference. In A. Duranti & C. Goodwin (Eds.), *Rethinking context: Language as an interactive phenomenon* (pp. 43-76). Cambridge, NY: Cambridge University Press.

Henderson, L. M., Yoder, P. J., Yale, M. E., & McDuffie, A. (2002). Getting the point: Electrophysiological correlates of protodeclarative pointing. *International Journal of Developmental Neuroscience*, *20*(3), 449-458.

Hottenroth, P.-M. (1982). The system of local deixis in Spanish. In J. Weissenborn, & W. Klein (Eds.), *Here and there: Cross-linguistic studies on deixis and demonstration* (pp. 133-153). Amsterdam: John Benjamins.

Jungbluth, K. (2003). Deictics in the conversational dyad: Findings in Spanish and some cross-linguistic outlines. In F. Lenz (Ed.), *Deictic conceptualisation of space, time and person* (pp. 13-40). Amsterdam: John Benjamins.

Kendon, A. (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University

Press.

Kita, S. (2003). *Pointing. Where language, culture, and cognition meet.* Hillsdale, NJ: Erlbaum.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and language*, *48*(1), 16-32.

Küntay, A., & Özyürek, A. (2002). Joint attention and the development of the use of demonstrative pronouns in Turkish. In B. Skarabela, S. Fish, & A. H. Do (Eds.), *Proceedings of the 26th annual Boston University Conference on Language Development* (pp. 336-347). Somerville, MA: Cascadilla Press.

Küntay, A., & Özyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal? *Journal of Child Language*, *33*, 303-320.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203-205.

Lakoff, R. (1974). Remarks on this and that. In M. W. La Galy, R. A. Fox, & A. Bruck (Eds.), *Papers from the tenth regional meeting: Chicago Linguistic Society* (pp. 345-356). Chicago.

Leavens, D. A., Hopkins, W. D., & Bard, K. A. (1996). Indexical and referential pointing in chimpanzees (Pan troglodytes). *Journal of Comparative Psychology*, *110*(4), 346.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: Bradford.

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. C. (1994). Deixis. In R. E. Asher (Ed.), *Encyclopedia of language and linguistics* (pp. 853-857). Oxford: Pergamon Press.

Levinson, S. C. (2004). Deixis. In L. Horn (Ed.), *The handbook of pragmatics* (pp. 97-121). Oxford: Blackwell.

Lyons, J. (1977). *Semantics. Volume 2*. Cambridge: Cambridge University Press.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Özyürek, A. (1998). An analysis of the basic meaning of Turkish demonstratives in face-to-face conversational interaction. In S. Santi, I. Guaitella, C. Cave, & G. Konopczynski (Eds.), *Oralité et gestualité: Communication multimodale, interaction: actes du colloque ORAGE 98* (pp. 609-614). Paris: L'Harmattan.

Peirce, C. S. (1931). *The collected writings of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Piwek, P., Beun, R. J., & Cremers, A. (2008). 'Proximal' and 'distal' in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics*, *40*, 694-718.

Quine, W. V. O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Rauh, G. (1983). Aspects of deixis. In G. Rauh (Ed.), *Essays on deixis* (pp. 9-60). Tübingen: Narr.

Russell, B. (1940). *An inquiry into meaning and truth*. London: George Allen & Unwin Ltd.

Senft, G. (2004). *Deixis and Demonstratives in Oceanic Languages*. Canberra: Pacific Linguistics.

Sherzer, J. (1973). Verbal and nonverbal deixis: The pointed lip gesture among the San Blas Cuna. *Language in Society*, *2*(01), 117-131.

Stevens, J., & Zhang, Y. (2013). Relative distance and gaze in the use of entity-referring spatial demonstratives: An event-related potential study. *Journal of Neurolinguistics*, *26*, 31-45.

Stevens, J., & Zhang, Y. (2014). Brain mechanisms for processing co-speech gesture: A cross-language study of spatial demonstratives. *Journal of Neurolinguistics*, *30*, 27-47.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child development*, *78*(3), 705-722.

Wilkins, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet* (pp. 171-215). Hillsdale NJ: Erlbaum.

Wittgenstein, L. (1953). *Philosophical Investigations.* Oxford, UK: Blackwell.

Chapter 2

The Interplay between Joint Attention, Physical Proximity, and Pointing Gesture in Demonstrative Choice: Evidence from Dutch and Turkish

Based on: Peeters, D., Azar, Z., & Özyürek, A. (2014). The interplay between joint attention, physical proximity, and pointing gesture in demonstrative choice. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 1144-1149). Austin, TX : Cognitive Science Society.

and

Peeters, D., Azar, Z., & Özyürek, A. (under review). The interplay between joint attention, physical proximity, and pointing gesture in demonstrative choice: Evidence from Dutch and Turkish.

Abstract

A fundamental property of language is that it allows establishing a joint focus of attention on a referent, for instance by the use of spatial demonstratives. Traditional accounts of demonstrative choice focus mainly on the physical proximity of the referent to the speaker. However, recent cross-linguistic and corpus based work, taking into account the multimodal and social context in which demonstratives are used, shows that such accounts are insufficient. Using a controlled elicitation task, we here tested the differential roles of the visual joint attention between speaker and addressee to a referent, the physical proximity of the referent to speaker and addressee, and the use of a pointing gesture in demonstrative choice in Dutch (which has a 2-term demonstrative system) and Turkish (which has a 3-term system). We also compared how these different factors played a role in the choice of demonstrative versus other referring expressions (i.e., in/definite terms). We found that in both languages 'proximal' demonstratives were used to refer to objects nearby the speaker and 'distal' demonstratives were used for referents not nearby the speaker. In Turkish, the distal term was also used when the referent was in the addressee's focus of visual attention. Pointing gestures were closely tied to the use of demonstratives but not to other (in)definite expressions. These findings confirm recent observational findings in showing that demonstrative choice goes beyond taking into account the physical proximity of a referent and is dependent on a subtle interplay between different context-dependent factors that can be employed in different ways by speakers of different languages.

The Interplay between Joint Attention, Physical Proximity, and Pointing Gesture in Demonstrative Choice: Evidence from Dutch and Turkish

> On Tuesday, when it hails and snows,
> The feeling on me grows and grows
> That hardly anybody knows
> If those are these or these are those.
>
> (A.A. Milne)

Establishing triadic joint attention to a referent is a very basic human communicative ability (Carpenter, Nagell, & Tomasello, 1998). A common way to do so is by using demonstrative pronouns and determiners (henceforth: demonstratives) such as *this* and *that* in English, often combined with a manual pointing gesture and a shift of gaze to the referent (Bakeman & Adamson, 1984; Clark & Bangerter, 2004; Diessel, 1999). Most languages contain more than one demonstrative term (Diessel, 2005), which implies that one has to choose among different options when using a demonstrative. Traditional accounts of demonstrative reference have generally been spatialist in nature, taking the relative proximity of a referent to the speaker as a fundamental criterion in demonstrative choice (e.g., Anderson & Keenan, 1985; Halliday & Hasan, 1977; Lyons, 1977, inter alia). Such egocentric proximity accounts argue that one selects a 'proximal' demonstrative to refer to an object that is physically relatively close to oneself, and a 'distal' demonstrative to refer to an object that is relatively far away from oneself. This speaker-centered view is omnipresent in reference grammars (Diessel, 2005) and in recent experimental studies (e.g., Coventry, Valdés, Castillo, & Guijarro-Fuentes, 2008; Stevens & Zhang, 2013).

However, evidence is accumulating showing that traditional accounts, based on physical proximity alone, are insufficient in explaining demonstrative choice, in particular when one considers the multimodal and visual context in which exophoric demonstrative use canonically occurs (e.g.,

Coventry, Griffiths, & Hamilton, 2014; Enfield, 2003; Hanks, 1990; Jones, 1995; Strauss, 2002). Descriptions of demonstrative systems in different languages from different language families (e.g., Dutch, English, Jahai, Jordanian Arabic, Turkish) for instance suggest that factors related to the locus of visual attention of the addressee and/or the cognitive (rather than physical) accessibility of a referent may play an important role in which demonstrative a speaker chooses (Burenhult, 2003; Jarbou, 2010; Küntay & Özyürek, 2006; Piwek, Beun, & Cremers, 2008; Stevens & Zhang, 2013). Küntay and Özyürek (2006), for example, suggest that the Turkish "medial" demonstrative *şu* is primarily used to shift the visual attention of the addressee when she does not (yet) visually attend to a certain referent, independent of the relative physical distance of that referent. Conversely, the "distal" demonstrative *o* was used for referents that were already in the addressee's focus of attention. As such, joint attention between speaker and addressee focused on a referent is not only often the desired outcome of exophoric demonstrative use (Diessel, 1999), but the presence or absence of joint attention to a referent at the moment a referring expression is instantiated may also drive the choice for a particular demonstrative over another. Being able to monitor and follow the gaze of an interlocutor is indeed a pivotal communicative skill and is often a prerequisite for successful communication (e.g., Bakeman & Adamson, 1984; Küntay & Özyürek, 2006; Senju & Csibra, 2008).

Another fundamental aspect of demonstratives in exophoric use is that they are canonically combined with a concomitantly produced manual pointing gesture, often in addition to other ostensive cues in the speaker's eye gaze and/or head and body orientation. An interesting observation is that in their contextual uses some demonstrative terms have been found to be more often combined with a manual pointing gesture than others (Burenhult, 2003; Küntay & Özyürek, 2006; Piwek et al., 2008; Senft, 2004). For instance, in a study by Piwek et al. (2008) native speakers of Dutch always produced a manual pointing gesture when uttering a proximal demonstrative, but not always when using a distal demonstrative term, and similar observations have been reported for native speakers of English

(Cooperrider, 2011, p. 73). It is an open question to what extent these findings interact with or are driven by other factors such as the relative proximity of the referents to the speaker (e.g., Bangerter, 2004).

In general, findings from relatively naturalistic settings can be tested and further specified in a controlled environment (Hanks, 2009). In the current study, going beyond previous experimental approaches to demonstrative choice (e.g., Coventry et al., 2008), we therefore orthogonally contrasted in a controlled setting three variables that may influence demonstrative choice: the visual joint attention between speaker and addressee to a referent prior to demonstrative use, the location of the referent and as such its physical proximity to the speaker, and the presence of a manual pointing gesture. We focus on Dutch (Study 1) and Turkish (Study 2), two languages with typologically different demonstrative systems (two-term versus three-term respectively). Previous observational work suggests that, beyond the physical proximity of a referent to the speaker, contextual factors related to the locus of visual attention of the addressee and the concurrent production of a pointing gesture also influence demonstrative choice in these two languages (see below). In contrast, if speakers of all languages apply an egocentric coordinate system and use demonstratives as a function of the referent's proximity to the speaker (Diessel, 2014), then the only difference between Dutch and Turkish should be that two-term Dutch makes a two-way distance contrast (close vs. far referents) whereas three-term Turkish makes a three-way contrast (for referents close to, at middle distance, or far from the speaker).

Finally, the current study not only focuses on the choice of demonstratives but also looks at noun phrases containing (in)definite articles in Dutch and definite nouns in Turkish (see below for examples). Demonstratives can be placed within a wider class of referring expressions (e.g., Ariel, 1988) and it is an open question whether similar factors influence the choice of demonstratives and the choice of definite and indefinite expressions in exophoric reference more broadly. Furthermore, pointing gestures may be paired with demonstratives more than with other referring expressions, because one function of demonstratives may be to direct the addressee's gaze to the gesture (Bangerter, 2004; Bühler, 1934).

Identification of the factors influencing the choice of referring expression is not only theoretically interesting, but may also inform computational models of reference production (see Van Deemter, Gatt, Van Gompel, & Krahmer, 2012).

Study 1

In Study 1 we tested demonstrative choice in Dutch. In a controlled elicitation task, we presented participants with visual scenes and asked them to complete a sentence from the perspective of the "speaker" in the scene. Three factors were orthogonally manipulated in these materials. First, the visual focus of attention of the addressee (and as such the joint attention between speaker and addressee to a referent before demonstrative choice) in the visual scenes was varied. Second, we varied the location of the referent and as such its physical distance from the speaker and addressee. A third factor manipulated was the presence or absence of an index-finger pointing gesture produced by the speaker.

Traditional (egocentric) proximity-based theories do not predict a difference in demonstrative choice based on the attentional focus of the addressee. In contrast, there are some preliminary indications that suggest that joint attention may play a role in demonstrative choice in Dutch. Piwek et al. (2008) had pairs of Dutch participants, consisting of an instructor and a builder, construct a small building using Lego blocks. Participants' speech, recorded during the task, was analyzed off-line and related to their focus of attention on a referent. A referent was coded as in the focus of attention, i.e. cognitively relatively accessible, when it was mentioned in the preceding utterance and/or when it was in an area toward which the speaker had explicitly directed the attention of the addressee already. It was found that participants used proximal demonstratives (*dit, deze* in Dutch) to refer to objects that were not in the focus of attention ('low cognitive accessibility') and distal demonstratives (*dat, die*) to objects that were in the focus of attention ('high cognitive accessibility').

On the basis of their results, Piwek et al. (2008) argue against proximity-based views of demonstrative choice. However, the physical proximity of the referents was not quantified. It is therefore unclear whether the physical proximity of referents (to speaker and/or to addressee) also influenced demonstrative choice and whether this interacted with the addressee's focus of attention. Furthermore, in their operationalization of cognitive accessibility the addressee's visual and cognitive foci of attention were collapsed. If the findings by Piwek et al. (2008) generalize to situations where the visual focus of attention of the addressee is manipulated, this would predict the use of proximal demonstratives for referents that are not in the addressee's visual focus of attention. In such a situation, a proximal demonstrative in Dutch would indeed allow strong indicating to shift the addressee's attention towards the referent (Piwek et al., 2008). Under such an account, Dutch speakers use distal demonstratives for referents that are already in the focus of visual attention of the addressee.

Piwek et al. (2008) thus argue that the relative physical distance of a referent does not primarily drive demonstrative choice. In contrast, egocentric proximity accounts predict that proximal demonstratives are used for referents close to the speaker and that the use of distal demonstratives increases with an increase in the relative physical distance of the referent from the speaker (e.g., Anderson & Keenan, 1985; Halliday & Hasan, 1977; Lyons, 1977). Our manipulation allows contrasting and testing these views. Moreover, egocentric proximity accounts do not predict an influence of the presence or absence of a pointing gesture on demonstrative choice. However, observational research suggests that within a language pointing gestures may be more closely tied to some demonstratives than to others (Burenhult, 2003; Cooperrider, 2011; Küntay & Özyürek, 2006; Piwek et al., 2008; Senft, 2004). We here further explore whether the presence of a manual pointing gesture is more closely tied to one demonstrative than to another in Dutch.

Finally, in the current paradigm participants were free in their choice of referring expression, which elicited not only demonstratives but also definite and indefinite articles. The advantage of this

approach compared to a forced choice paradigm is that it does not give away the aim of the study to the participants, and as such decreases the chances that the results reflect their meta-linguistic intuitions about demonstratives. Similar to English, Dutch speakers may use not only 'proximal' demonstratives (*dit/deze*, "this", depending on the gender of the noun corresponding to the referent) or 'distal' demonstratives (*dat/die,* "that"), but also definite articles (as in *de bal,* "the ball") or indefinite articles (as in *een bal*, "a ball") in reference to an object. This allowed testing whether the same contextual factors influence demonstrative choice and the choice of other referring expressions more broadly. For instance, it is not unlikely that speakers would use more definite than indefinite articles in Dutch when objects are nearby and/or when their addressee is already visually attending to a referent.

Method

Participants

Twenty native speakers of Dutch studying in Nijmegen (13 female; mean age 22.2) participated in return for payment. They had normal or corrected-to-normal vision and no history of language impairment.

Materials

The materials consisted of 64 target triplets of still images that contained a speaker, an addressee, and an object. Each triplet consisted of an introductory picture, a target picture, and a concluding picture. Figure 1 shows an example of one such triplet.

Figure 1. *Each trial consisted of a sequence of three pictures (introductory picture, target picture, and concluding picture; converted to grayscale). Participants read out loud and completed the sentence presented in the target picture.*

Three independent variables were orthogonally manipulated in the target pictures. First, the location of the object was either close to the speaker, close to the addressee, at middle distance from both speaker and addressee, or relatively far away from speaker and addressee. Second, there was either visual joint attention or no visual joint attention between speaker and addressee to the object before the referential expression was to be used. In the case of no visual joint attention, the speaker looked at the referent object while the addressee looked at another part of the visual scene (e.g., a painting, see Fig.1). Third, the speaker either produced a pointing gesture towards the object or not. It is not uncommon in natural interactions for speakers to refer to an object using speech and gesture while their addressee is not yet looking at them or their referent (e.g., Küntay & Özyürek, 2006). The three manipulated factors are henceforth called Location, Joint Attention, and Pointing Gesture respectively. Figure 2 shows a subset of the target pictures.

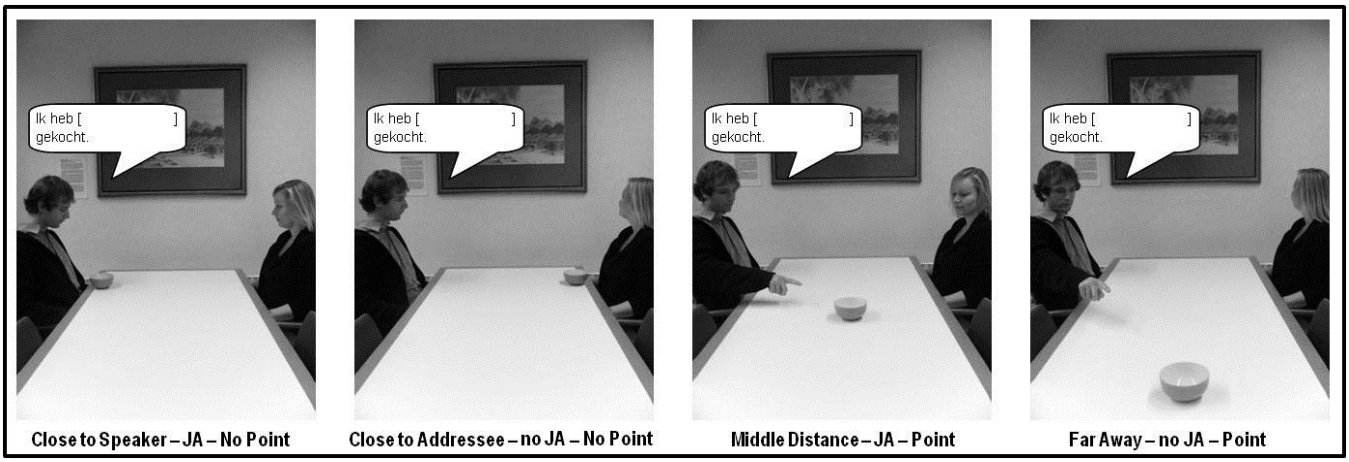

Figure 2. *Subset of target pictures used in the elicitation task in Dutch (converted to grayscale). Below the pictures it is indicated to which condition the picture belonged. The object could be in four different locations. There could be joint attention (JA) or no joint attention (no JA) between speaker and addressee to the object, and the speaker could either manually point towards the object or not.*

In all target pictures the speaker and addressee were located opposite each other. The referent object was a ball, a bowl, a bucket, or a cup (n=16 each). Whether the person on the left or the right in the picture was the speaker was counterbalanced and indicated by a text balloon close to the mouth of the speaker. The rationale behind showing the pictures from a sagittal viewpoint (instead of from behind the speaker) was that in this way the exact focus of visual attention of the speaker and addressee was clear to the participants. Every target picture contained a text balloon presenting a declarative sentence that was missing a referring expression and could be completed using a demonstrative or an article and a noun. All sentences were of the form 'Ik heb + [ ] + verb form' ('I have + [ ] + verb form'), which elicited sentences such as 'Ik heb *dit* kopje schoongemaakt' (I have cleaned *this* cup) or 'Ik heb *die* bal meegebracht' (I have brought *that* ball). The verb form was specified in the sentences; only the referring expression was left out.

The introductory picture was always the same as the target picture except for the presence of the text balloon, and served to introduce the visual context of each trial. In the concluding picture there was always joint attention between speaker and addressee to the referent object. As such every referential act successfully resulted in joint attention to the referent.

In addition to the 64 experimental trials, 64 triplets were created that were used as fillers in the task. The rationale behind adding the filler trials was to ensure that participants would not guess that the study targeted demonstrative choice. The filler images depicted a single speaker and one object. A large number of different objects were used. In these filler trials, the text balloons contained a sentence that did not elicit a demonstrative or article (e.g. the Dutch translation of 'What a nice [ ]').

Procedure

Participants were tested individually in a sound-proof booth. The experiment was presented on a computer screen using *Presentation* software (Neurobehavioral Systems). Participants were instructed to carefully look at the pictures, and read aloud and complete the sentence shown in the picture that was presented in the center of the screen. Trials were presented one by one. Each trial started with an introductory picture at the left part of the screen. After 1500 ms, the target picture appeared in the center of the screen, while the introductory picture remained visible. After having read aloud and completed the written sentence in the target picture, participants pressed the spacebar on a keyboard, which resulted in presentation of the concluding picture. Together with the other two pictures, this picture remained visible on the screen for another 1500 ms, after which the next trial started. The 128 trials were presented in a fully randomized order, different for each participant.

Data Coding

Throughout the task, participants' speech was continuously recorded by a voice recorder (Olympus Imaging Corp.) linked to an external microphone. The elicitation task yielded 128 uttered sentences per participant. Speech was transcribed off-line by a native speaker of Dutch and each referring expression was coded for the presence of a demonstrative ('proximal' *dit/deze* or 'distal' *dat/die*) or an article (definite *de/het* or indefinite *een*) preceding the noun. In line with previous studies (e.g. Piwek et al., 2008), demonstratives were collapsed across grammatical gender in the analysis.

Results

Participants produced a demonstrative or article on 95.3% of all trials. In this dataset, 32.4% of trials contained a demonstrative (9.0% '*dit/deze'* + 23.4% '*dat/die'*) and 67.6% of trials contained an article (20.2% definite + 47.4% indefinite). Figure 3 gives an overview of the proportion of use of each of these terms for each level of each of the three independent variables.

A multinomial logistic regression analysis was carried out on the use of the proximal and distal demonstrative and the definite and indefinite article, with the three independent variables (Joint Attention, Location, and Pointing Gesture) as predictors (forced entry) and their interaction terms included in the model. Because indefinite articles were used more often than the three other types of referring expression and because their use was relatively stable across the levels of each independent variable (see Figure 3), they were selected as the reference category. Interaction terms were eliminated backwards from the model to test whether this changed the fit of the model to the data. The results showed that the final model, consisting of the three predictors, explained significantly more variance than the baseline model, χ^2 (15) = 295.3, $p < .001$; R^2 = .22 (Cox & Snell), .24 (Nagelkerke). This final model contained no interaction terms, because removal of each interaction term did not significantly

alter how well the model explained the variance in the data. There was no sign of overdispersion [Pearson χ^2 (30) = 20.1, p = .914; Deviance χ^2 (30) = 20.8, p = .895].

Table 1 presents the results of the analysis. There are three main findings. First, the presence or absence of joint attention between speaker and addressee to a referent did not predict the use of referring expression (i.e. any demonstrative or article) in Dutch. Second, the location of the referent significantly predicted the use of demonstratives (vs. indefinite articles), but not the use of definite articles (vs. indefinite articles). The odds ratio and Figure 3 show that, globally, proximal demonstratives were used for referents close to the speaker and not for referents at the other three locations, and vice versa for distal demonstratives. A change in referent location from middle distance to far away also significantly decreased the use of a proximal demonstrative. Third, the presence of a pointing gesture significantly predicted the use of both demonstrative terms (vs. indefinite articles), but not the use of definite articles (vs. indefinite articles). Demonstratives were used more often when there was a pointing gesture compared to when there was no pointing gesture.

Discussion

The first study investigated three factors that might influence how we refer to entities in the world around us by zooming in on the choice of demonstratives and articles in Dutch.

We found that Dutch speakers took into account the physical location of the referent in their demonstrative choice. Globally, referents nearby the speaker elicited a proximal demonstrative whereas referents in three physically more distant regions elicited a distal demonstrative. Interestingly, no linear increase of distal demonstrative use was found as a function of an increase in relative physical distance from the speaker to the referent. Rather, speakers seemed to differentiate between a zone close to the speaker and the rest of the extra-linguistic space (Enfield, 2003). In addition, the presence of a pointing

gesture elicited the use of demonstrative terms rather than the use of (in)definite articles. Arguably, in cases when a distal demonstrative is produced, a pointing gesture could narrow down the addressee's search space in looking for the intended referent. Piwek et al. (2008) suggested that the Dutch distal demonstrative may be used systematically in cases when no strong indicating is necessary because the referent is already in the focus of attention. In contrast with this suggestion, we did not observe an influence of the presence or absence of the addressee's visual attention to a referent on demonstrative (or article) choice.

In sum, the location of the referent and the presence of a manual pointing gesture are exploited in the use of different demonstratives in Dutch. These factors do not necessarily play a similar role in influencing the choice of definite and indefinite articles. These findings will be further discussed in the General Discussion.

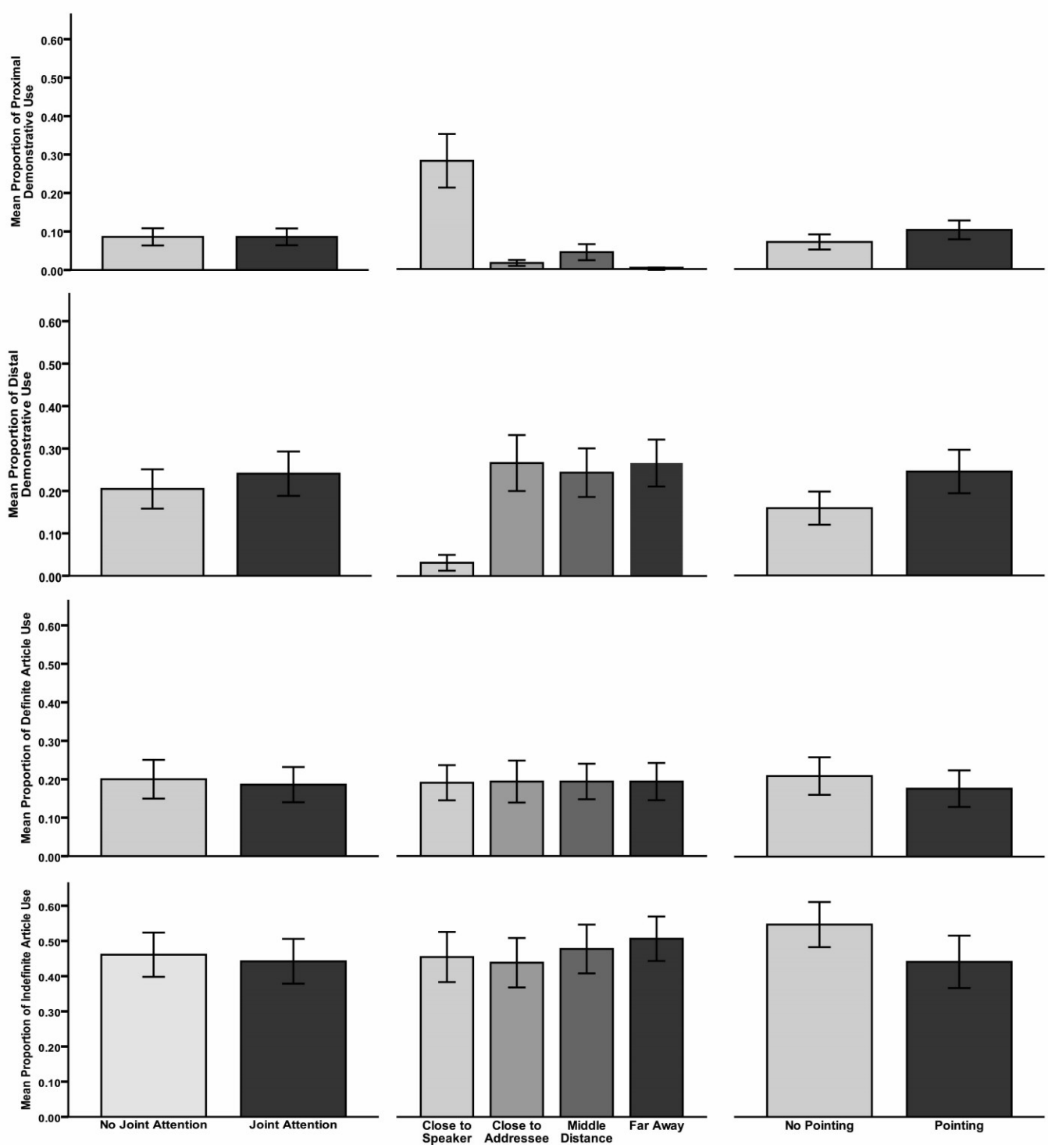

Figure 3. *Separate panels for the mean proportion of use of 'proximal' demonstratives, 'distal' demonstratives, definite articles, and indefinite articles in Study 1 (Dutch) as a function of the three factors manipulated in the target pictures (Joint Attention, Location, and Pointing Gesture). Error bars represent standard errors of the mean.*

Table 1. *Outcome of the multinomial logistic regression analysis on the referring expressions elicited in Dutch. The use of the indefinite article was used as the baseline category.*

				95% CI for Odds Ratio		
		B (SE)	Wald χ^2 (df)	Lower	Odds	Upper
Proximal Demonstrative (*dit/deze*)						
Intercept		-4.77 (1.02)***	22.09 (1)			
Joint Attention	(no vs. yes)	-0.03 (0.23)	0.02 (1)	0.62	0.97	1.52
Location	(close to Spkr)	4.60 (1.01)***	20.65 (1)	13.70	99.64	724.93
	(close to Addr)	1.76 (1.10)	2.54 (1)	0.67	5.80	50.27
	(mid-distance)	2.70 (1.04)**	6.73 (1)	1.93	14.90	114.79
	(far away)					
Pointing Gesture	(no vs. yes)	-0.55 (0.23)*	5.77 (1)	0.37	0.58	0.90
Distal Demonstrative (*dat/die*)						
Intercept		-0.90 (0.17)	0.30 (1)			
Joint Attention	(no vs. yes)	-0.21 (0.15)	2.02 (1)	0.60	0.81	1.08
Location	(close to Spkr)	-2.05 (0.34)***	36.07 (1)	0.07	0.13	0.25
	(close to Addr)	0.15 (0.19)	0.62 (1)	0.80	1.16	1.68
	(mid-distance)	-0.03 (0.19)	0.02 (1)	0.67	0.98	1.42
	(far away)					
Pointing Gesture (no vs. yes)		-0.67 (0.15)***	19.42 (1)	0.38	0.51	0.69
Definite Article (*de/het*)						
Intercept		-0.91***	23.36 (1)			
Joint Attention	(no vs. yes)	0.03	0.04 (1)	0.77	1.03	1.39
Location	(close to Spkr)	0.09	0.18 (1)	0.72	1.10	1.67
	(close to Addr)	0.15	0.46 (1)	0.76	1.16	1.76
	(mid-distance)	0.06	0.08 (1)	0.70	1.06	1.61
	(far away)					
Pointing Gesture	(no vs. yes)	-0.05	0.09 (1)	0.71	0.96	1.29

*p <.05, ** $p \leq .01$, *** $p < .001$

Study 2

The first study showed that demonstrative choice in Dutch, a language with a simple two-term demonstrative system, was dependent on (at least) two contextual factors, i.e. the physical location of the referent and the presence of a pointing gesture. This raises the question whether these findings are language-specific or, instead, generalize to other languages with typologically different demonstrative systems. In a second step, therefore, we investigated whether the same three contextual factors influenced demonstrative choice in Turkish, a language with a three-term demonstrative system.

Traditionally, the Turkish demonstrative system has been described both as a proximity-based system and as a person-oriented system. In addition to a 'proximal' demonstrative (*bu*) that is used for referents relatively close to the speaker and a 'distal' demonstrative (*o*) that is used for referents relatively distant from speaker and addressee, these accounts differ in their description of the 'medial' demonstrative term (*şu*). The (egocentric) proximity-based account argues that this term is used exophorically for referents that are somewhat removed from the speaker (Kornfilt, 1997), whereas the person-oriented account claims that it is used for referents in physical proximity to the addressee (Lyons, 1977). Both these accounts thus explain Turkish demonstrative choice in terms of physical proximity; the latter only includes the addressee as a zero-point from where proximity may be determined. In contrast, as outlined in the Introduction, more recent observational findings suggest that the medial term (*şu*) is used independent of the distance of the referent to the speaker, as a function of whether the attention of the addressee is already on the referent or not. If the referent is not yet in focus of visual attention of the addressee, *şu* will be used to shift the addressee's attention towards the referent (Özyürek, 1998; Küntay & Özyürek, 2006). Conversely, in addition to referring to objects that are relatively far away, the distal term *o* may be used for referents in the addressee's focus of attention (Küntay & Özyürek, 2006). Besides testing whether and to what extent the current findings on Dutch

generalize to a language with a typologically different demonstrative system, our manipulation allowed contrasting and testing these three different views in a controlled setting.

Method

Participants

Twenty native speakers of Turkish studying in Istanbul (15 female; mean age 21.4) participated in the study and received course credits in return for their participation. They had normal or corrected-to-normal vision and no history of language impairment.

Materials

The images used in Study 2 were identical to the images used in Study 1, except for two changes. First, the Dutch speakers in the pictures were replaced by Turkish speakers. Second, the target sentences were Turkish equivalents of the Dutch sentences used in Study 1. All target sentences were of the form '[ ] +ben + verb form' (in English: '[ ] +I + verb form'), which elicited sentences such as '*Bu* bardağı ben temizledim' (I have cleaned *this* cup) or '*O* topu ben getirdim' (I have brought *that* ball). The verb form was always specified in the sentences; only the referring expression was left out. Filler sentences were used that would not elicit a demonstrative (e.g. the Turkish translation of 'What a nice [ ]').

Procedure and Data Coding

The procedure was identical to Study 1. For Study 2, the speech was transcribed off-line and coded by a native speaker of Turkish. Data coding was similar to Study 1. Turkish has no article system, but to indicate the specificity of the referent Turkish speakers may use accusative marking on direct

objects (see Göksel & Kerslake, 2005). In our data, definite noun phrases with an accusative case marker were commonly used (without a demonstrative), henceforth referred to as 'definite nouns'. An example sentence is as follows:

'Top-u ben getir-di-m'

Ball-ACC I bring-PAST-1st PERSON

"I brought the ball"

Results

Participants produced a demonstrative or definite noun on 95.5% of all trials. In this dataset, 64.7% of trials contained a demonstrative (26.4% *bu* + 17.6% *şu* + 20.8% *o*) and 35.3% of trials contained a definite noun. Figure 4 gives an overview of the proportion of use of each of these terms for each level of each of the three independent variables.

A multinomial logistic regression analysis was carried out on the use of the three types of demonstrative and the definite noun, with the three independent variables (Joint Attention, Location, and Pointing Gesture) as predictors (forced entry) and their interaction terms included in the model. Because definite nouns were used more often than the three other types of referring expression and because their use was relatively stable across the levels of each independent variable (see Figure 4), they were selected as the reference category. Interaction terms were eliminated backwards from the model to test whether this changed the fit of the model to the data. The results showed that the final model, consisting of the three predictors, explained significantly more variance than the baseline model, χ^2 (15) = 385.1, p < .001; R^2 = .27 (Cox & Snell), .29 (Nagelkerke). This final model contained no interaction terms, because removal of each interaction term did not significantly alter how well the model explained the

variance in the data. There was no sign of overdispersion [Pearson χ^2 (30) = 14.9, p = .990; Deviance χ^2 (30) = 15.8, p = .984].

Table 2 presents the results of the analysis. Again, there are three main findings. First, the presence or absence of joint attention between speaker and addressee to a referent significantly predicted the use of the distal demonstrative (*o*) in Turkish. When there was joint attention to the referent compared to no joint attention, the use of *o* (rather than a definite expression) increased significantly. Second, the location of the referent significantly predicted the type of demonstrative that was used. The proximal term (*bu*) was mainly used for referents close to the speaker. In addition, a change in referent location from close to the speaker to close to the addressee, and from close to the addressee to middle distance, led to a significant increase in the use of the medial term (*şu*). The distal term (*o*) was used more when the referent was close to the addressee compared to when it was close to the speaker. In addition, it was used more for referents far away compared to referents at middle distance. Third, the presence of a pointing gesture also significantly predicted the use of the demonstratives *bu* and *şu*, rather than the use of a definite noun.

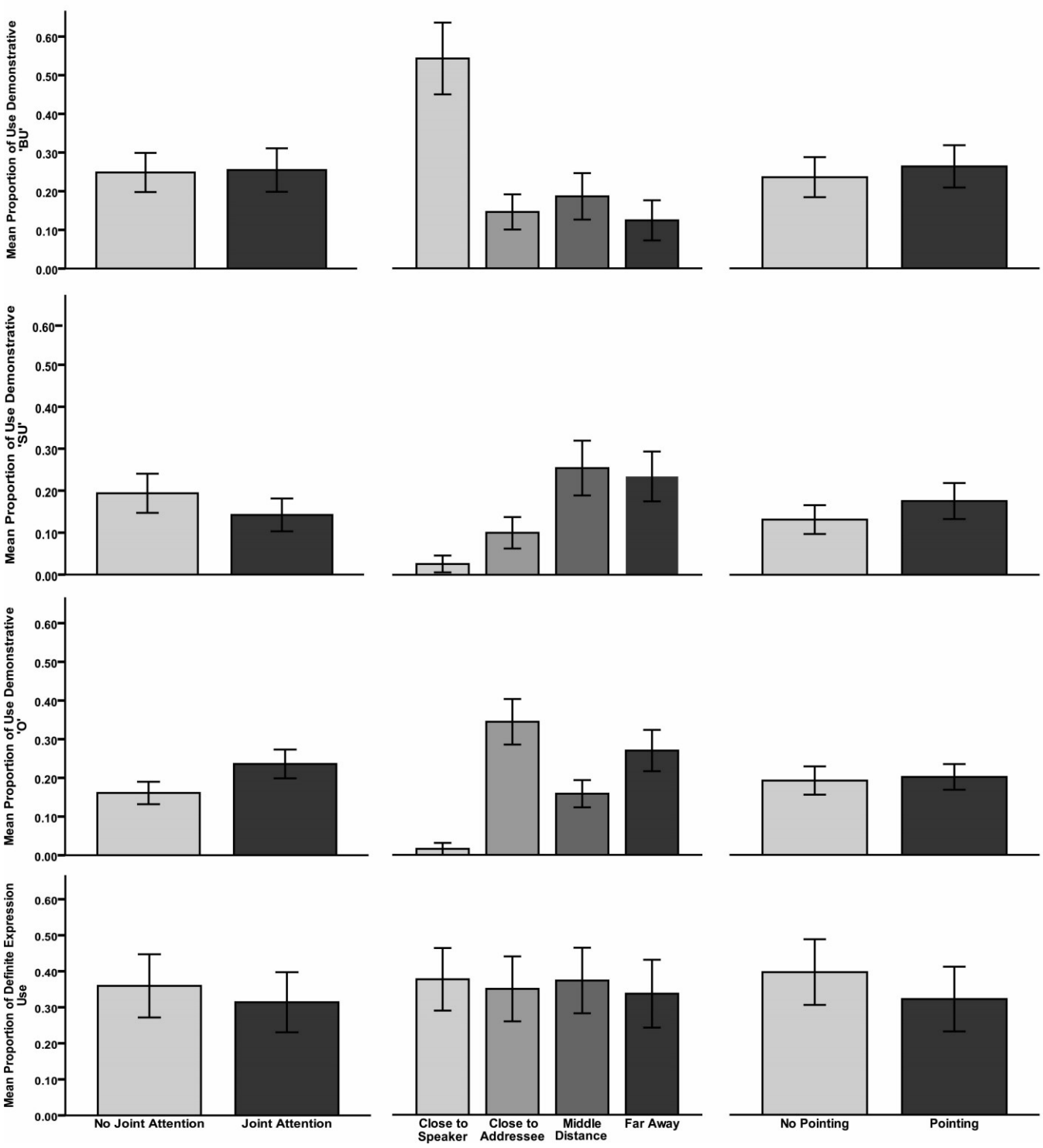

Figure 4. *Separate panels for the mean proportion of use of the three Turkish demonstrative types (bu, şu, and o), and the use of definite nouns in Study 2 (Turkish) as a function of the three factors manipulated in the target pictures (Joint Attention, Location, and Pointing Gesture). Error bars represent standard errors of the mean.*

Table 2. *Outcome of the multinomial logistic regression analysis on the referring expressions elicited in Turkish. The use of a definite expression was used as the baseline category.*

				95% CI for Odds Ratio		
		B (SE)	Wald χ^2 (df)	Lower	Odds	Upper
Proximal Demonstrative (*bu*)						
Intercept		-0.69**	10.10 (1)			
Joint Attention	(no vs. yes)	-0.16	1.02 (1)	0.63	0.86	1.16
Location	(close to Spkr)	1.36***	37.17 (1)	2.52	3.89	6.03
	(close to Addr)	0.13	0.24 (1)	0.69	1.13	1.88
	(mid-distance)	0.30	1.53 (1)	0.84	1.36	2.20
	(far away)					
Pointing Gesture	(no vs. yes)	-0.30*	3.90 (1)	0.55	0.74	1.00
Medial Demonstrative (*şu*)						
Intercept		-0.06	0.08 (1)			
Joint Attention	(no vs. yes)	0.19	1.22 (1)	0.86	1.21	1.71
Location	(close to Spkr)	-2.33***	38.04 (1)	0.05	0.10	0.20
	(close to Addr)	-0.90***	13.22 (1)	0.25	0.41	0.66
	(mid-distance)	-0.03	0.02 (1)	0.65	0.97	1.46
	(far away)					
Pointing Gesture	(no vs. yes)	-0.53**	9.32 (1)	0.42	0.59	0.83
Distal Demonstrative (*o*)						
Intercept		0.24	1.69 (1)			
Joint Attention	(no vs. yes)	-0.54**	10.57 (1)	0.42	0.58	0.81
Location	(close to Spkr)	-2.97***	38.26 (1)	0.02	0.05	0.13
	(close to Addr)	0.22	1.16 (1)	0.84	1.24	1.84
	(mid-distance)	-0.64**	8.05 (1)	0.34	0.53	0.82
	(far away)					
Pointing Gesture	(no vs. yes)	-0.27	2.72 (1)	0.55	0.76	1.05

*p <.05, ** $p \leq .01$, *** $p < .001$.

Discussion

A second study was carried out to investigate to what extent three contextual factors would influence demonstrative choice in Turkish, a language with a three-term demonstrative system. The results of Study 2 critically showed that, in Turkish, the 'distal' term *o* was used when the addressee's focus of visual attention was already on the referent. This pattern thus falsifies person-oriented and egocentric proximity-based accounts of Turkish demonstrative reference (Kornfilt, 1997; Lyons, 1977), as it is in line with a more recent proposal that suggests a crucial role for the focus of visual attention of the addressee in demonstrative choice in Turkish (Küntay & Özyürek, 2006).

Regardless of the focus of visual attention of the addressee, and in line with previous descriptions, the 'proximal' term *bu* was used to refer to objects near the speaker. Interestingly, the 'medial' term *şu* was used more when the referent was at middle distance and far away compared to when it was close to speaker or addressee (see Figure 4). Although this effect showed up in the manipulation of the location of the referent, it may have been driven by the attentional status of the addressee. When in everyday life a referent is outside of the shared space between speaker and addressee, it may be assumed to be more likely outside of their attentional focus as well, eliciting *şu* (cf. Küntay & Özyürek, 2006). Again, proximity-based accounts of demonstrative choice cannot explain such findings, in this case because the referent close to the addressee was equally far from the speaker as the referent in middle distance.

Furthermore, the results of the second study showed that the specific interplay between the referent's location and the addressee's attentional focus may be specific to demonstrative terms, and not to other definite expressions. Finally, similar to the findings in Dutch, pointing gestures were found to be tied more closely to demonstratives than to other definite expressions.

We will compare the Turkish results to the Dutch results and further discuss the theoretical implications of our findings in the General Discussion, presented next.

General Discussion

A fundamental property of language is that it allows us to refer to entities in the world around us (e.g., Clark, 1996; Tomasello, 2008). The current study found that the visual attentional focus of one's addressee, the physical location of the referent, and the concomitant use of a pointing gesture may differentially influence how speakers do this in two languages (Dutch and Turkish) that have typologically different demonstrative systems (two-term versus three-term). It further specifies that the influence of these factors may play a significant role in demonstrative choice, but not necessarily in the choice of (in)definite articles in Dutch and definite expressions more broadly in Turkish. Earlier accounts of demonstrative reference have sometimes explained demonstrative choice in terms of a single factor such as physical proximity (e.g., Anderson & Keenan, 1985; Halliday & Hasan, 1977; Lyons, 1977) or a general notion of 'accessibility' (e.g., Jarbou, 2010; Piwek et al., 2008). Here we show that different visual contextual factors may play differential roles within and across demonstrative systems and that speakers have different strategies at their disposal. Furthermore, our findings confirm the significance of orthogonally contrasting different factors in the same study (cf. Coventry et al., 2014; Stevens & Zhang, 2013), the importance of testing observational findings in a controlled paradigm (Hanks, 2009), and the added value of cross-linguistically comparing typologically different demonstrative systems using the same methodological approach.

The Dutch two-term system and the Turkish three-term system showed several similarities in patterns of demonstrative use in a triadic situation. The 'proximal' demonstratives

dit/deze in Dutch and *bu* in Turkish were used in a similar, speaker-anchored way in both languages to differentiate a zone near the speaker from the rest of the extra-linguistic space. The 'distal' demonstratives *dat/die* in Dutch and *o* in Turkish were used for referents not near the speaker. Furthermore, in both Dutch and Turkish no influence of the attention of the addressee and the physical location of the referent was found on the choice of referring expressions that did not contain a demonstrative.

As a language-specific strategy, the 'distal' term *o* in Turkish was used when the referent was already in the focus of attention of the addressee. This finding confirmed previous observational results (Küntay & Özyürek, 2006) and falsified traditional accounts of the Turkish demonstrative system, which were purely based on the physical proximity of the referent to the speaker and addressee (Lyons, 1977; Kornfilt, 1997). Interestingly, in this respect Turkish differs from other demonstrative systems (see Kirsner & Van Heuven, 1988). For instance, in Jordanian Arabic it is the proximal demonstrative that is used for entities with high perceptibility to the addressee (Jarbou, 2010). Interestingly, research on other languages has also found an influence of the addressee's attention on demonstrative choice (e.g., Burenhult, 2003; Jarbou, 2010; Piwek et al., 2008). Together, these findings confirm that joint attention may not only be the aim and result of using a referring expression (Diessel, 1999), but also a driving force in demonstrative choice.

Another finding specific to Turkish was the higher use of the 'medial' demonstrative *şu* for objects away from speaker and addressee compared to objects close to one of the two interlocutors. We tentatively suggested that this effect of 'location' may be driven by a strong positive correlation in everyday life between the physical proximity of a referent to speaker and addressee and the likelihood of that referent being in their focus of visual attention. Note that this

finding is in sharp contrast with what was predicted by the person-oriented account of Turkish demonstrative choice, namely that it is mainly used for referents close to the addressee (Lyons, 1977).

Our findings also have implications for more general accounts of demonstrative reference. It is surprising that, especially in the linguistic literature since Bühler (1934), demonstrative systems have often been described in egocentric terms (e.g., Anderson & Keenan, 1985; Halliday & Hasan, 1977; Lyons, 1977; see also Russell, 1940). Even recently, Diessel (2014) for instance stated that, when using a demonstrative term, 'speakers of all languages employ an egocentric coordinate system that is anchored by the speaker's body at the time of the utterance (p.128)'. Such a view is largely based on linguistic intuitions (cf. Enfield 2003) and not on careful empirical testing or in-depth observational analysis of patterns of use in context. There are now many empirical reasons to believe that demonstrative use is driven by sociocentric rather than egocentric motives (e.g., Enfield, 2003; Hanks, 1990; Jungbluth, 2003; Laury, 1996; Piwek et al., 2008). The current study confirms such a sociocentric view in showing that speakers may take into account the addressee's focus of attention in their choice of a demonstrative. Even the finding that the 'proximal' demonstrative in both Dutch and Turkish was used in a speaker-anchored fashion does not imply that this was done with egocentric intentions or from an egocentric coordinate system. One may use a 'proximal' demonstrative in a particular context simply to inform one's addressee that a referent is close to oneself, taking into account both oneself and the addressee. After all, the addressee is just as necessary and prominent in a communicative act as the speaker (Jones, 1995), and establishing reference is often a joint, collaborative act in a social context (Clark, 1996; Clark, Schreuder, & Buttrick, 1983; Hanks, 1990).

Our results further suggest that people may divide space into a zone near the speaker, possibly a zone near the addressee, and the rest of the physical space. Indeed, it is not uncommon that physical space is transformed into meaningful space in everyday interactions (Enfield, 2003; Kendon, 1977; Scheflen & Ashcraft, 1976). How the space is exactly carved up in the mind of interlocutors may strongly depend on the context. In certain situations, unlike in the current study, a distinction will be made between the shared space in between speaker and addressee in a conversational dyad, and the rest of the space outside of the dyad (Jungbluth, 2003; see also Chapter 3 of this thesis). This may have consequences for speakers' demonstrative choice, in that all referents within the dyad may elicit a proximal demonstrative term (Jungbluth, 2003). The current results confirm that in demonstrative choice the context-dependent division of space is more important than purely physical factors such as the relative proximity of a referent to the speaker (Enfield, 2003).

Both in the current Dutch and Turkish study, we found that the presence of a pointing gesture was more closely tied to the use of demonstratives than to the use of (in)definite expressions that did not contain a demonstrative. Bangerter (2004, p. 418) suggested that demonstratives may direct gaze to a concurrently used pointing gesture when the gesture carries the main informational burden (see also Bühler, 1934). Our results are in line with this idea and we suggest that a pointing gesture may demarcate the addressee's search space in cases when a distal demonstrative is used in reference to an object not near the speaker. Previous research has shown that speakers design the exact kinematic properties of their index-finger pointing gesture, such as its velocity, trajectory, and the duration of its post-stroke hold-phase, by taking into account the mental state of their addressee (Cleret de Langavant et al., 2011; see also Chapter 4

of the current thesis). A demonstrative could then indeed be used to make the addressee pay attention to such an effort.

To conclude, the current study showed differential roles of visual joint attention, physical proximity, and the presence of pointing gestures in demonstrative choice in two languages with typologically different demonstrative systems. Our findings confirm that the very basic human communicative ability of establishing triadic joint attention to a referent turns out to be dependent on a subtle interplay between different context-dependent factors, as reflected in one's choice of demonstrative. These results open up new avenues towards understanding the complex interplay between contextual factors involved in demonstrative choice within and across languages.

References

Anderson, S. R., & Keenan, E. L., (1985). Deixis. In T. Shopen (Ed.), *Language typology and syntactic description* (pp. 259-308). Cambridge: Cambridge University Press.

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, *24*, 65-87.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development*, 55, 1278-1289.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*(6), 415-419.

Bühler, K. (1934). *Sprachtheorie*. Jena: Fischer.

Burenhult, N. (2003). Attention, accessibility, and the addressee: The case of the Jahai demonstrative ton. *Pragmatics, 13*, 363-379.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 255, Vol. 63,* 1-174.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Bangerter, A. (2004). Changing ideas about reference. In I. A. Noveck, & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 25-49). Basingstoke: Palgrave Macmillan.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *Journal of verbal learning and verbal behavior*, *22*(2), 245-258.

Cleret de Langavant, L., Remy, P., Trinkler, I., McIntyre, J., Dupoux, E., Berthoz, A., & Bachoud-Lévi, A. C. (2011). Behavioral and neural correlates of communication via pointing. *PloS one*, *6*(3), e17719.

Cooperrider, K. (2011). Reference in action: Links between pointing and language. Doctoral dissertation, University of California, San Diego.

Coventry, K. R., Griffiths, D., & Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: Describing and remembering object location. *Cognitive psychology*, *69*, 46-70.

Coventry, K. R., Valdés, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language within your reach: Near–far perceptual space and spatial demonstratives. *Cognition*, *108*, 889-895.

Diessel, H. (1999). *Demonstratives. Form, Function, and Grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2005). Distance contrasts in demonstratives. In M. Haspelmat, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The World Atlas of Language Structures* (pp. 170-173). Oxford: Oxford University Press.

Diessel, H. (2014). Demonstratives, Frames of Reference, and Semantic Universals of Space. *Language and Linguistics Compass*, *8*(3), 116-132.

Enfield, N. J. (2003). Demonstratives in space and interaction: Data from Lao speakers and implications for semantic analysis. *Language*, 82-117.

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. New York: Routledge.

Halliday, M. A. K., & Hasan, R. (1977). *Cohesion in English*. London: Longman Group Ltd.

Hanks, W. F. (1990). *Referential practice: Language and lived space among the Maya*. Chicago: University of Chicago Press.

Hanks, W. F. (2009). Fieldwork on deixis. *Journal of Pragmatics, 41,* 10-24.

Jarbou, S. O. (2010). Accessibility vs. physical proximity: An analysis of exophoric demonstrative practice in Spoken Jordanian Arabic. *Journal of Pragmatics*, *42*(11), 3078-3097.

Jones, P. (1995). Philosophical and theoretical issues in the study of deixis: A critique of the standard account. In K. Green (Ed.), *New essays in Deixis: Discourse, narrative, literature* (pp. 27-48). Amsterdam: Rodopi.

Jungbluth, K. (2003). Deictics in the conversational dyad: Findings in Spanish and some cross-linguistic outlines. In F. Lenz (Ed.), *Deictic conceptualisation of space, time and person* (pp. 13-40). Amsterdam: John Benjamins.

Kendon, A. (1977). Spatial organization in social encounters: The F-formation system. In A. Kendon (Ed.), *Studies in the behavior of social interaction* (pp. 179-208). Lisse: Peter de Ridder Press.

Kirsner, R. S., & Van Heuven, V. J. (1988). The significance of demonstrative position in modern Dutch. *Lingua*, *76*(2), 209-248.

Kornfilt, J. (1997). *Turkish*. London: Routledge.

Küntay, A., & Özyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal? *Journal of Child Language*, *33*, 303-320.

Laury, R. (1996). Conversational use and basic meaning of Finnish demonstratives. In A. E. Goldberg (Ed.), *Conceptual structure, discourse and language* (pp. 303-319). Stanford: CSLI publications.

Levinson, S. C. (2000). Presumptive meanings: The theory of generalized conversational implicature. Cambridge, MA: MIT Press.

Lyons, J. (1977). *Semantics. Volume 2*. Cambridge: Cambridge University Press.

Özyürek, A. (1998). An analysis of the basic meaning of Turkish demonstratives in face-to-

face conversational interaction. In S. Santi, I. Guaitella, C. Cave, & G. Konopczynski (Eds.), *Oralité et Gestualité: Communication multimodale, interaction* (pp. 609-614). Paris: L'Harmattan.

Piwek, P., Beun, R. J., & Cremers, A. (2008). 'Proximal' and 'distal' in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics*, *40*, 694-718.

Russell, B. (1940). *An inquiry into meaning and truth*. London: George Allen & Unwin Ltd.

Scheflen, A. E., & Ashcraft, N. (1976). *Human territories. How we behave in space-time.* Englewood Cliffs, NJ: Prentice-Hall Inc.

Senft, G. (2004). *Deixis and Demonstratives in Oceanic Languages*. Canberra: Pacific Linguistics.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, *18*(9), 668-671.

Stevens, J., & Zhang, Y. (2013). Relative distance and gaze in the use of entity-referring spatial demonstratives: An event-related potential study. *Journal of Neurolinguistics*, *26*, 31-45.

Strauss, S. (2002). This, that, and it in spoken American English: a demonstrative system of gradient focus. *Language Sciences*, *24*(2), 131-152.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Van Deemter, K., Gatt, A., van Gompel, R. P. G., & Krahmer, E. (2012). Toward a computational psycholinguistics of reference production. *Topics in cognitive science*, *4*(2), 166-183.

Chapter 3

Electrophysiological Evidence for the Role of Shared Space in Online Comprehension of Spatial Demonstratives

Based on: Peeters, D., Hagoort, P., & Özyürek, A. (2015). Electrophysiological evidence for the role of shared space in online comprehension of spatial demonstratives. *Cognition*, 136, 64-84. doi:10.1016/j.cognition.2014.10.010.

Abstract

A fundamental property of language is that it can be used to refer to entities in the extra-linguistic physical context of a conversation in order to establish a joint focus of attention on a referent. Typological and psycholinguistic work across a wide range of languages has put forward at least two different theoretical views on demonstrative reference. Here we contrasted and tested these two accounts by investigating the electrophysiological brain activity underlying the construction of indexical meaning in comprehension. In two EEG experiments, participants watched pictures of a speaker who referred to one of two objects using speech and an index-finger pointing gesture. In contrast with separately collected native speakers' linguistic intuitions, N400 effects showed a preference for a proximal demonstrative when speaker and addressee were in a face-to-face orientation and all possible referents were located in the shared space between them, irrespective of the physical proximity of the referent to the speaker. These findings reject egocentric proximity-based accounts of demonstrative reference, support a sociocentric approach to deixis, suggest that interlocutors construe a shared space during conversation, and imply that the psychological proximity of a referent may be more important than its physical proximity.

Electrophysiological Evidence for the Role of Shared Space in Online Comprehension of Spatial Demonstratives

> "Distance cannot be what distinguishes the meanings of these two demonstratives"
> (Enfield, 2003, p. 104)

A fundamental property of language is that it can be used to refer to entities in the extra-linguistic physical context of a conversation. Using a spatial demonstrative such as "this" or "that" in combination with a pointing gesture allows one to establish a joint focus of attention on a referent (Diessel, 2006), and to directly anchor one's communication to the material world (Clark, 2003; Weissenborn & Klein, 1982). The production and comprehension of such referring expressions entails a complex interaction between language, gesture, space, and attention. This has long been a topic of interest in not only linguistics (e.g., Bühler, 1934; Jakobson, 1971; Jespersen, 1922; Levinson, 1983; Rommetveit, 1968) but also in anthropology (see Hanks, 1990; 2005), cognitive science (e.g., Kemmerer, 1999), and philosophy (e.g., Peirce, 1931; Russell, 1940). Furthermore, establishing and understanding triadic reference is a milestone in the acquisition of language and social skills (Bakeman & Adamson, 1984; Butterworth, 2003; Carpenter, Nagell, & Tomasello, 1998; Clark & Sengul, 1978; Tomasello, Carpenter, & Liszkowski, 2007). Here we focus on the neural and cognitive mechanisms underlying the comprehension of referential speech acts that contain a spatial demonstrative and a concurrently produced pointing gesture. The present study investigates this issue by contrasting and testing two different theoretical views on demonstrative reference. To set the stage for the description of the present study, we will first outline these two views, which were put forward on the basis of

previous typological and experimental studies on the production and comprehension of such spatial demonstratives.

Demonstrative systems in the world's languages have long mainly been described in egocentric and speaker-anchored terms. Speakers would use demonstrative pronouns and determiners on the basis of the proximity of the referent from their own physical location (cf. Anderson & Keenan, 1985; Diessel, 2005; Fillmore, 1982; Halliday & Hasan, 1977; Hottenroth, 1982; Lakoff, 1974; Lyons, 1977; Rauh, 1983) and as such do not consider demonstrative reference a joint activity but rather an egocentric act. In a survey of reference grammars and other typological sources on 234 languages from a large range of different language families and geographical areas, Diessel (1999, 2005) found that the large majority (i.e., 227) of languages in his corpus used a proximal-distal contrast in the description of their adnominal demonstrative system. In English, for instance, the adnominal demonstratives "this" and "that" would be used comparably to the adverbials "here" and "there", in that a distance scale is used with the speaker as its deictic center (Anderson & Keenan, 1985; Clark & Sengul, 1978; Diessel, 1999; Fillmore, 1982; Halliday & Hasan, 1977; Lakoff, 1974; Lyons, 1977). When the two demonstratives are used contrastively, "this" is used for referents in relative proximity to the deictic center (i.e., the speaker), whereas "that" is used for referents relatively remote from the speaker's location (Diessel, 2005; Lakoff, 1974; Levinson, 1983). Also three-term systems, such as Spanish, have been described as speaker-anchored systems, with a "medial" demonstrative that would be used for entities at medial distance from the speaker (Anderson & Keenan, 1985; Diessel, 1999; Fillmore, 1982; Levinson, 2004). Such an egocentric speaker-anchored explanation of deictic reference is found not only in linguistic typology, but also in philosophical approaches to indexicality (e.g., Russell, 1940). We will call this first theoretical view *the egocentric proximity*

account. Recently, this account has been taken as a starting point in psycholinguistic experimental work.

Coventry et al. (2008), for instance, experimentally investigated the production of English and Spanish demonstratives in a controlled "memory game" paradigm. Participants were seated at a table with twelve spatial locations marked by colored dots in front of them at different distances. They were instructed to name objects that were placed on the different dots by using a three-term structure (e.g., *that red triangle*). When objects were within the participant's reach, both English and Spanish participants preferred to use a proximal demonstrative (*this* in English; *este* in Spanish). When the objects were placed outside of the participants' reach, a non-proximal demonstrative was preferred (*that* in English; *ese* or *aquel* in Spanish). The authors conclude that spatial demonstratives are used on the basis of a perceptual distinction between peripersonal space near the speaker and extrapersonal space far away from the speaker (but see Bonfiglioli, Finocchiaro, Gesierich, Rositani, & Vescovi, 2009; Kemmerer, 1999).

Recently, Stevens and Zhang (2013) studied the comprehension of demonstratives by showing participants (native speakers of English) images of a speaker, an addressee, and several potential referents. In all the images, the speaker pointed and gazed toward a referent, which was always a blue cat. In a 2x3 design, the locus of attention of the addressee (on the referent, not on the referent), and the location of the referent (in speaker-associated space, i.e., close to speaker, in hearer-associated space, i.e., close to addressee, or in non-associated space, i.e., remote from both) were manipulated. Each image was then combined once with the auditory expression "this one" and once with the expression "that one". Participants looked at the images and listened to the referential utterances, and were asked to judge by pressing one of two buttons whether the audiovisual scene (*this one* or *that one* in combination with the image) was semantically

congruent or incongruent. In addition, their electroencephalogram (EEG) was recorded continuously during the experiment.

Stevens and Zhang found that participants judged a proximal demonstrative (*this*) congruent for referents in speaker-associated space, and a distal demonstrative (*that*) congruent for referents in hearer-associated and non-associated space. Thus, the participants' linguistic intuitions as expressed in their congruency judgments were in line with the egocentric proximity account outlined above. In addition, it was found that reaction times were faster for congruent demonstrative usages, but only when speaker and hearer shared gaze at the referent. Thus, the egocentric proximity account would only hold for situations in which speaker and addressee both already focus their gaze on the referent. The event-related potential (ERP) data further complicated the story, in only showing an effect in one of the six conditions. When the referent was in hearer-associated space, and there was shared gaze between speaker and hearer on the referent, a late negative deflection was found for the proximal demonstrative compared to the distal demonstrative. The authors name this an N600 effect, which would be a delayed variant of the canonical N400 effect for semantically anomalous words (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980).

The egocentric proximity account has been challenged by a number of studies that argued that a referent's relative proximity to the speaker is not the only or not the most important parameter in determining the selection of demonstratives in a particular context (Anderson & Keenan, 1985; Burenhult, 2003; Da Milano, 2007; Himmelmann, 1996; Küntay & Özyürek, 2006; Levinson, 1983; Piwek, Beun, & Cremers, 2008). Some three-term systems would have the middle demonstrative refer to entities close to the addressee (e.g., Japanese: Diessel, 1999, 2005; Levinson, 2004) thus taking the addressee's location into account. In other languages, the

visibility, geographical location (e.g., uphill or downhill), and height of the referent (Anderson & Keenan, 1985; Diessel, 1999), or the presence or absence of joint attention between speaker and addressee (Küntay & Özyürek, 2006; Levinson, 2004) on the referent would also play a role in which demonstrative a speaker would select. Nevertheless, the egocentric proximity account is omnipresent in typologies of demonstrative systems (Diessel, 2005).

A theoretical alternative to the egocentric proximity account was put forward by Jungbluth (2003) in line with Weinrich (1988) and Laury (1996). This second theoretical view has been termed *the dyad-oriented account*. On the basis of fieldwork observations on Spanish, Jungbluth (2003) proposes that the physical orientation of speaker and addressee relative to each other plays a crucial role in explaining which demonstrative a Spanish speaker would select to use. More specifically, when speaker and addressee are face-to-face, as in a conversational dyad, a proximal demonstrative (*este* in Spanish) would be used to refer to an object within the shared space between speaker and addressee, irrespective of whether the referent is close to the speaker or not. In other words, every referent inside the conversational dyad is treated as proximal without any further differentiation (Jungbluth, 2003, p.19). Referents outside of the dyad, for instance when located behind the addressee, would elicit a distal demonstrative (*aquel* in Spanish*)*. Thus, this view on demonstratives moves away from the egocentric, 'monadic' typologies that underline a referent's relative physical proximity to the speaker, and supports a more sociocentric, dyadic view of multimodal reference in which the spatial locations of speaker, addressee, and referent play an important role (Weinrich, 1988). This sociocentric approach to deixis is in line with other work arguing that the egocentric speaker-anchored proximity view is too simplistic because it does not acknowledge that establishing reference is a social, interactive

process that always takes place in a socio-cultural framework (Enfield, 2003; Hanks, 1990; 2005).

In the current study we will contrast and empirically test these two theoretical accounts on spatial demonstratives, which were mainly based on the study of demonstrative production, in two electrophysiological experiments in demonstrative comprehension.

The present study

The first aim of the present study is to contrast and test the two theoretical accounts outlined above by recording participants' electroencephalogram (EEG) and time-lock event-related potentials (ERPs) to the onset of an auditorily presented demonstrative. In order to create a visual, referential context, we presented participants with pictures in which there were two objects and a speaker. One object was always close to the speaker and another at a larger relative distance from the speaker, and on each trial the speaker referred to one of the two objects by manually pointing at it and using a demonstrative (*this* or *that*) embedded in a sentence. The two objects were located in one of two differentially oriented axes (i.e., lateral vs. sagittal) which allowed us to contrast the two theoretical views. Figure 1 below shows a subset of the stimuli.

The egocentric proximity account (e.g., Anderson & Keenan, 1985; Clark & Sengul, 1978; Diessel, 1999; Fillmore, 1982; Halliday & Hasan, 1977; Lakoff, 1974; Lyons, 1977; Rauh, 1983; Russell, 1940) predicts that perceiving a distal demonstrative ("that") in reference to an object close to the speaker should be reflected in a higher processing cost compared to perceiving a proximal demonstrative ("this") in reference to the same object. Similarly, perceiving a proximal demonstrative in reference to an object that is relatively remote from the speaker should also be experienced as caused by a violation. According to this account, it should

not matter whether objects are oriented on a sagittal or on a lateral plane between speaker and addressee. More specifically, this account thus predicts a Demonstrative x Distance interaction, and no Demonstrative x Orientation interaction.

In contrast, the dyad-oriented account (Jungbluth, 2003; Weinrich, 1988) predicts that people prefer a proximal demonstrative when speaker and addressee are oriented in a dyad and the referent is positioned in any location within the shared space in that dyad. This is the case for both referents in our sagitally oriented picture stimuli because the speaker and the participant addressee are sitting face-to-face, with the referents between them. It further predicts a preference for a proximal demonstrative for the object close to the speaker in the lateral orientation of objects, because this object may be considered to be inside the shared dyad space. Conversely, it predicts a preference for a distal demonstrative for the distal object in the lateral orientation, because this object is clearly outside of the dyad. In contrast with the egocentric account, this account thus predicts no Demonstrative x Distance interaction, but a Demonstrative x Orientation interaction effect instead.

The current study goes beyond the previous electrophysiological study by Stevens and Zhang (2013) on demonstrative comprehension in several ways. Firstly, it is the first study investigating the neural and cognitive mechanisms underlying demonstrative comprehension by an implicit addressee. It may be the case that overhearing a referential speech act uttered by a speaker for an addressee, as in Stevens and Zhang (2013), taps into different neural and cognitive mechanisms than being the (implicit) addressee of a referential speech act (e.g., Schober & Clark, 1989). Second, it is likely that people's linguistic intuitions about demonstratives do not align with their actual everyday usage and comprehension of demonstratives (Enfield, 2003; Piwek et al., 2008). Therefore, we will avoid using an experimental task that gives away the aim

of the study to the participants. In addition, we will use a large number of different referents throughout the experiment and embed the demonstratives in a wider sentence context, aspects which were both absent in Stevens and Zhang (2013).

Finally, Stevens and Zhang (2013) argue that the N600 effect they found is similar to the canonical N400 effect generally found to semantic violations on the noun. However, they did not include a condition in their experiment in which they compared the two types of violation. Friedrich and Friederici (2010) have shown that hearing an incorrect label to an object elicits a widespread semantic N400 effect on the noun. In the present study, we compare the time-course and topography of such an N400 effect to an effect of a "demonstrative violation" in the same study to the same visual materials.

Experiment 1

Method

Participants

Twenty-seven participants (mean age 20.9, range 18-27, three male) took part in the experiment. They were all right-handed as assessed by a Dutch translation of the Edinburgh Inventory for hand dominance (Oldfield, 1971). Data from four participants were discarded due to a large number of trials that contained eye blinks and movement artifacts. All participants were Dutch, studying in Nijmegen, and Dutch was their native language. They had normal or corrected-to-normal vision, no language or hearing impairments or history of neurological disease. They provided written informed consent and were paid for participation.

Materials and Pretest of Materials

The experimental materials consisted of 120 introductory pictures, 240 target pictures, and 240 spoken Dutch sentences. The pictures were selected after pretesting from an initial set of 140 introductory and 280 target pictures (see below). Half of the critical pictures were sagitally oriented and the other half of the pictures were laterally oriented (see Figure 1). In the laterally oriented pictures, the position of the speaker (left or right) was counterbalanced. All pictures contained two similar objects, one close to the speaker and one further away from the speaker. In total, 120 different object pairs were used. In half of the critical pictures, the speaker was looking and pointing at the proximate object (the *proximate* condition). In the other half, she was looking and pointing at the remote object (the *distal* condition). The size of the laterally oriented pictures was 15 x 9 cm and the sagitally oriented pictures 11 x 12 cm, subtending visual angles of respectively 7.8 x 4.7 and 5.7 x 6.2 degrees.

The 240 sentences were spoken by a female native speaker of Dutch and digitized at a sample frequency of 44.1 kHz. They consisted of 120 sentence pairs. The two sentences in each pair only differed in the demonstrative (proximal versus distal) they contained. All sentences started as follows: "*Ik heb* [demonstrative] [noun] *etc.*" (see Table 1 for examples translated into English) such that the words preceding the demonstrative were always the same across all conditions. For every sentence, the onsets of the demonstrative and the following noun were determined by using a speech analysis software package (Praat, version 5.2; www.praat.org). The sentences had an average duration of 2176 ms (*SD* = 224). The critical demonstratives had an average duration of 197 ms (*SD* = 67) and the critical nouns had an average duration of 409 ms (*SD* = 132). Sentences were equalized in maximum amplitude.

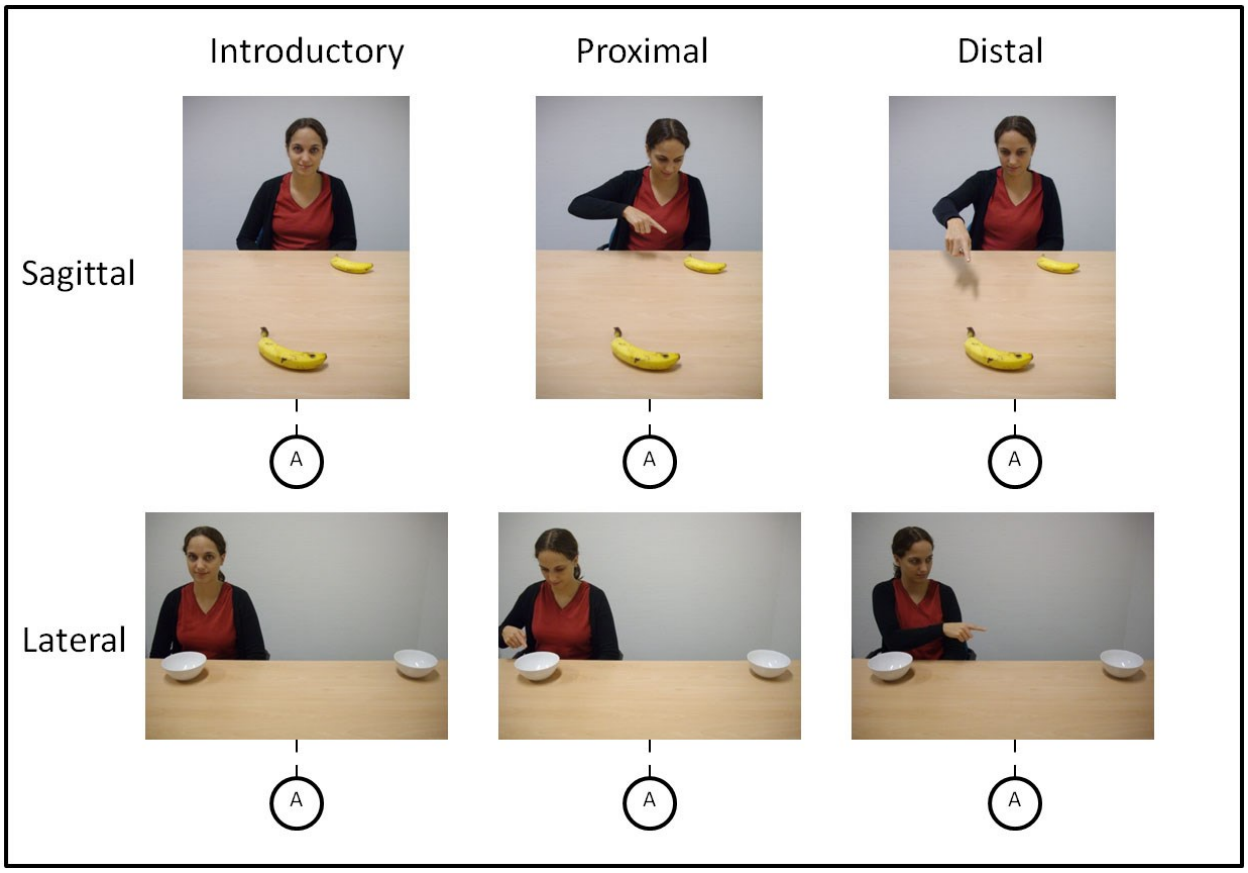

Figure 1. *Subset of stimuli used in Experiment 1. Each row contains an introductory picture, a target picture in which the speaker refers to the proximate object, and a target picture in which she refers to the distal object. The top row shows pictures with a sagittal orientation of objects. The bottom row shows pictures with a lateral orientation of objects. The position of the participant (i.e. the implicit Addressee) is marked below the pictures.*

The experiment contained a demonstrative manipulation and a noun manipulation. The main independent variables in the demonstrative manipulation were Object Orientation (lateral or sagittal), Distance of the referent from the speaker (proximate or distal), and Demonstrative (proximal or distal). The demonstrative manipulation was done by having each picture paired with an auditorily presented sentence that could contain a proximal demonstrative (*dit* or *deze*, "this") or a distal demonstrative (*dat* or *die*, "that"). The experiment was carried out in Dutch, in

which generally a distinction is made between proximal (*dit, deze*) and distal (*dat, die*) demonstratives, with separate terms for neuter (*dit, dat*) and common gender (*deze, die*) nouns following the demonstrative.

The main independent variables in the noun manipulation were Object Orientation (lateral or sagittal), Distance of the referent from the speaker (proximate or distal), and Noun (congruent or incongruent). The noun manipulation was done by having each picture paired with an auditorily presented sentence that contained a correct noun-referent (e.g., "plate" for a plate) or an incorrect noun-referent (e.g., "mango" for a plate). In the noun manipulation, the demonstrative preceding the noun was always similar across noun congruency conditions. Table 1 shows an overview of the different manipulations.

To select the best materials, all pictures were presented in a pre-test to 16 native Dutch participants (mean age 20.4, range 18-26) who did not participate in the main experiment. They were shown the critical pictures, one by one, on a computer screen in a soundproof booth, and were asked to type the demonstrative and the noun that they would use imagining they were the speaker in the picture. This provided us with the demonstratives that participants intuitively prefer in different conditions and the labels that participants use for the depicted objects. Whenever the speaker pointed to the close object, participants used a proximal demonstrative ("this") in 97.86 % of the cases. When the speaker pointed to the remote object, participants used a distal demonstrative ("that") in 98.21 % of the cases. This preference was irrespective of the orientation (lateral vs. sagittal) of objects. From the 140 pictures, the 120 pictures that were labeled most consistently were selected for the main experiment.

Table 1. *Design of both Experiments 1 and 2 with examples of the auditory stimuli (translated from Dutch).*

Condition	Orientation	Distance	Example sentences	Picture
Proximal Demonstrative				
	Sagittal			
		Proximate	"I have bought **this** plate at the market "	Plate-close
		Distal	"I have bought **this** plate at the market"	Plate-far
	Lateral			
		Proximate	"I have found **this** vase in the cupboard"	Vase-close
		Distal	"I have found **this** vase in the cupboard"	Vase-far
Distal Demonstrative				
	Sagittal			
		Proximate	"I have bought **that** plate at the market"	Plate-close
		Distal	"I have bought **that** plate at the market"	Plate-far
	Lateral			
		Proximate	"I have found **that** vase in the cupboard"	Vase-close
		Distal	"I have found **that** vase in the cupboard"	Vase-far
Congruent Noun				
	Sagittal			
		Proximate	"I have bought this **plate** at the market"	Plate-close
		Distal	"I have bought that **plate** at the market"	Plate-far
	Lateral			
		Proximate	"I have found this **vase** in the cupboard"	Vase-close
		Distal	"I have found that **vase** in the cupboard"	Vase-far
Incongruent Noun				
	Sagittal			
		Proximate	"I have bought this **mango** at the market"	Plate-close
		Distal	"I have bought that **mango** at the market"	Plate-far
	Lateral			
		Proximate	"I have found this **spoon** in the cupboard"	Vase-close
		Distal	"I have found that **spoon** in the cupboard"	Vase-far

Procedure

Participants were seated in a comfortable chair in a dimly lit room at a distance of 110 cm from a computer screen. Pictures were presented on the screen using *Presentation* software (Neurobehavioral Systems) and speech was presented through EEG-compatible headphones.

Figure 2 shows the structure of a trial. A trial consisted of a fixation cross (200 ms), an introductory picture (800 ms), another fixation cross (200 ms), the target picture (2500 ms), and a symbol (- -) during which participants could blink their eyes (3000 ms). The spoken sentence was presented 500 ms after target picture onset and lasted, on average, 2176 ms. The introductory picture was presented to familiarize participants with the objects in the picture and to mark the speaker's referential, communicative intention. Directly looking at someone is often interpreted as an ostensive signal that establishes the speaker's intention to communicate (e.g., Csibra, 2010). As such, we hoped that participants would consider themselves the addressee of the speaker's utterance, as in a referential triadic context in which speakers alternate their gaze between the referent and their addressee (e.g., Bakeman & Adamson, 1984).

Participants were instructed to carefully look at the pictures and listen to the sentences. Their task was to push a button whenever the speaker in the pictures referred to both objects in the picture (*catch trials,* 10 % of all trials). When the speaker was referring to only one object, participants did not push a button (*target trials*, 90 % of all trials). They were instructed to blink their eyes only during the presentation of the symbol (- -). They were instructed to fixate on a central point on the screen and not to blink or make any saccades during the presentation of the target pictures. Participants correctly identified 89 % of all catch trials.

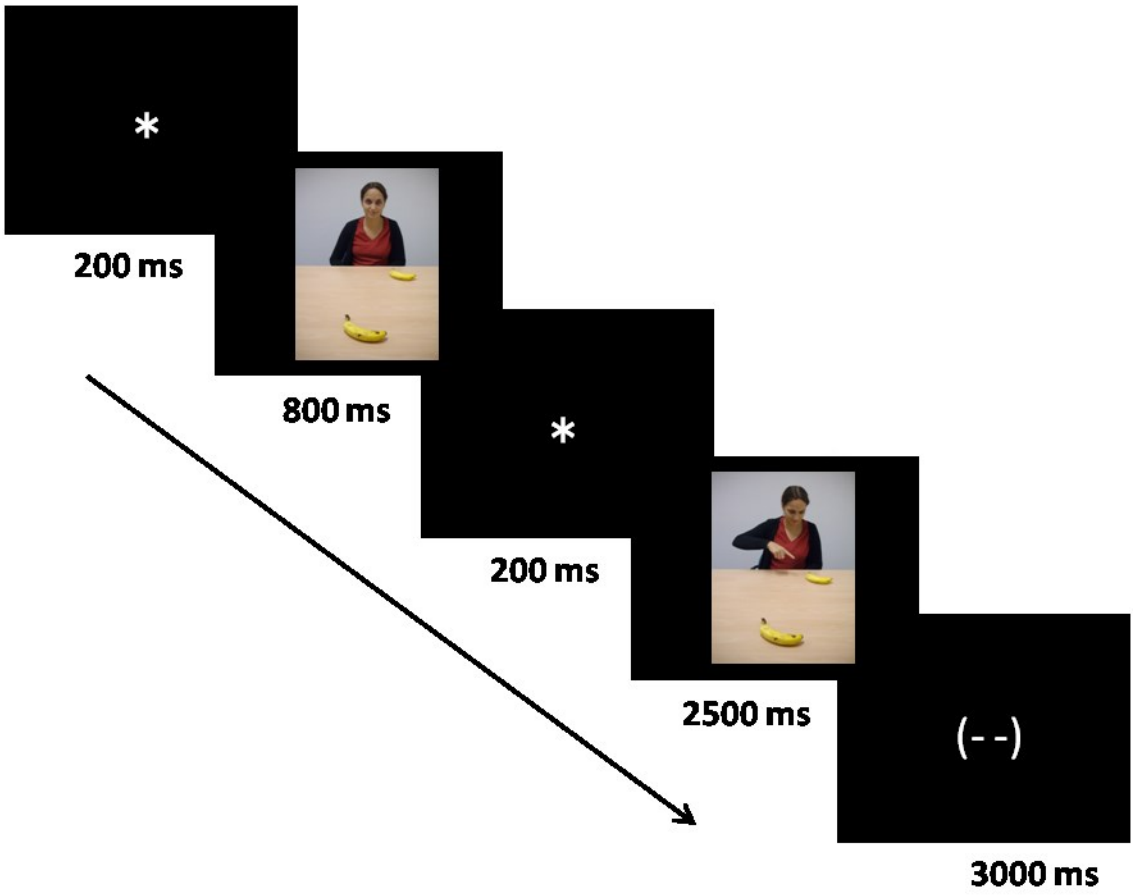

Figure 2. *Overview of the trial structure used in Experiments 1 and 2. During the presentation of the second picture a spoken sentence was presented.*

The experiment consisted of four blocks of 60 trials. A different randomized list was used for each participant. The experiment started with a practice session of twelve practice items, not used in the main experiment, to familiarize the subjects with the experiment. The experimental session lasted approximately 55 minutes.

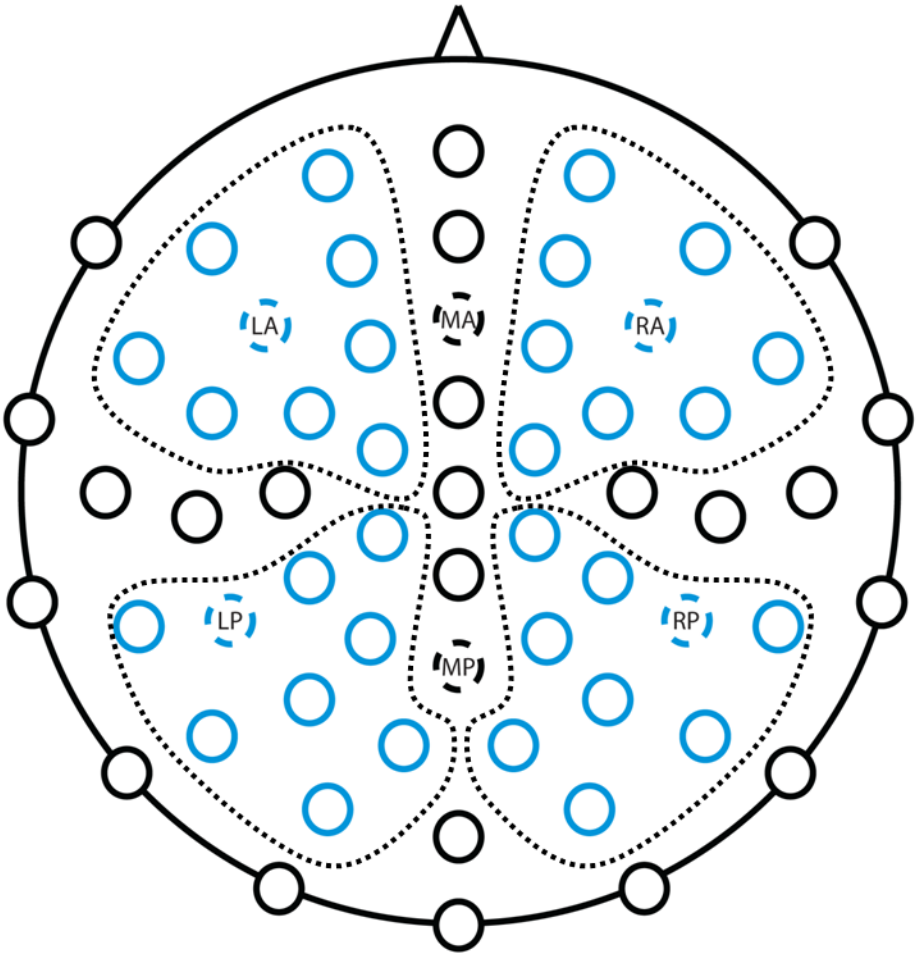

Figure 3. *Electrode montage. The four quadrants used in the analysis of the electrophysiological data (in addition to the vertical midline column) are circled and highlighted in blue. The electrode locations indicated with dashed lines refer to the electrode sites displayed in Figures 4-7 and 9-12 (left anterior, LA; middle anterior, MA; right anterior, RA; left posterior, LP; middle posterior, MP, right posterior, RP).*

EEG Recording and analysis

The electroencephalogram (EEG) was recorded continuously from 59 active electrodes held in place on the scalp by an elastic cap. Figure 3 shows the electrode montage. In addition to the 59 scalp sites, three external electrodes were attached to record EOG, one below the left eye

(to monitor for vertical eye movement/blinks), and two on the lateral canthi next to the left and right eye (to monitor for horizontal eye movements/saccades). Finally, one electrode was placed over the left mastoid bone and one over the right mastoid bone. All electrode impedances were kept below 20 KΩ. The continuous EEG was recorded with a sampling rate of 500 Hz, a low cut-off filter of 0.01 Hz and a high cut-off filter of 200 Hz. EEG was filtered offline (high-pass at 0.01 Hz and low-pass at 40 Hz). All electrode sites were online referenced to the electrode placed over the left mastoid and re-referenced offline to the average of the right and left mastoids.

Epochs from 200 ms preceding the onset of the demonstrative ("this" or "that") to 800 ms after the onset of the demonstrative, and epochs from 200 ms preceding the onset of the noun to 800 ms after the onset of the noun were selected using Brain Vision Analyzer software (Version 2.0, Brain Products, Munich). The 200 ms pre-stimulus period was used as a baseline. Trials containing ocular or muscular artifacts were not taken into consideration in the averaging process (16.8 % of all data). The mean amplitudes of the ERP waveforms for each condition per subject were entered into repeated measures ANOVAs. In the demonstrative manipulation, the independent variables were Object Orientation (lateral or sagittal; henceforth: Orientation), Distance of the referent from the speaker (proximate or distal; henceforth: Distance), Demonstrative (proximal or distal), Quadrant (left anterior, LA; right anterior, RA; left posterior, LP; right posterior, RP; vertical midline, VM), and Electrode. In the noun manipulation, the independent variables were Orientation (lateral or sagittal), Distance (proximate or distal), Noun (congruent or incongruent), Quadrant (LA, RA, LP, RP, VM), and Electrode. For both manipulations, we performed a time-window analysis for subsequent 100 ms time-windows in the 800 ms following noun onset (in the noun manipulation) and in the 800 ms following

demonstrative onset (in the demonstrative manipulation) respectively. The Geisser and Greenhouse (1959) correction was applied to all analyses with more than one degree of freedom in the numerator (corrected degrees of freedom and *p*-values are reported). Only significant results at the 5 % level are reported.

Results

Table 2 gives an overview of the results of the time-window analysis of the ERPs time-locked to the onset of the demonstrative. The overall analysis revealed most importantly a significant Demonstrative x Orientation interaction effect in time-windows between 100 and 600 ms after demonstrative onset. Follow-up analyses revealed a significant main effect of Demonstrative in these time-windows for the picture stimuli with a sagittal orientation of referents. Figure 4 shows this effect, which denotes a more negative wave for the distal demonstrative compared to the proximal demonstrative and is widespread over the scalp (i.e., it does not interact with Quadrant). Importantly, no interaction was found with Distance, implying that the effect was similar for the referent close to the speaker compared to the referent remote from the speaker. In contrast, no significant main effect of Demonstrative was found in any of the time-windows for the pictures with a lateral orientation of referents. Figure 5 shows the grand average waveforms and the topographic plots corresponding to the difference waves for different time-windows for the pictures with a lateral orientation of referents. Table 2 and Figure 6 show that the Demonstrative x Distance interaction predicted by the egocentric proximity account was not present in the data.

Table 2. *Results of the ERP analyses time-locked to the onset of the demonstrative in Experiment 1. This and subsequent tables present* F-*values and significance levels for separate analyses performed for 100 ms time-windows.*

	100-200	200-300	300-400	400-500	500-600	600-700	700-800
EXPERIMENT 1							
Overall							
Demonstrative x Orientation x Quadrant	1.49	1.91	1.19	< 1	< 1	1.31	< 1
Demonstrative x Orientation x Distance	2.56	< 1	< 1	< 1	< 1	< 1	< 1
Demonstrative x Orientation	13.91**	6.08*	5.48*	5.00*	5.45*	1.44	< 1
Demonstrative x Quadrant	< 1	< 1	4.11*	2.96	1.26	2.45	2.08
Demonstrative x Distance	1.22	< 1	< 1	< 1	< 1	< 1	4.05
Demonstrative	< 1	14.42**	1.30	2.55	4.91*	< 1	< 1
Lateral Orientation							
Demonstrative	2.14	< 1	< 1	< 1	< 1	1.15	< 1
Sagittal Orientation							
Demonstrative	11.67**	21.24***	4.61*	7.91*	9.70*	< 1	< 1

* $p < .05$; ** $p < .01$; *** $p < .001$

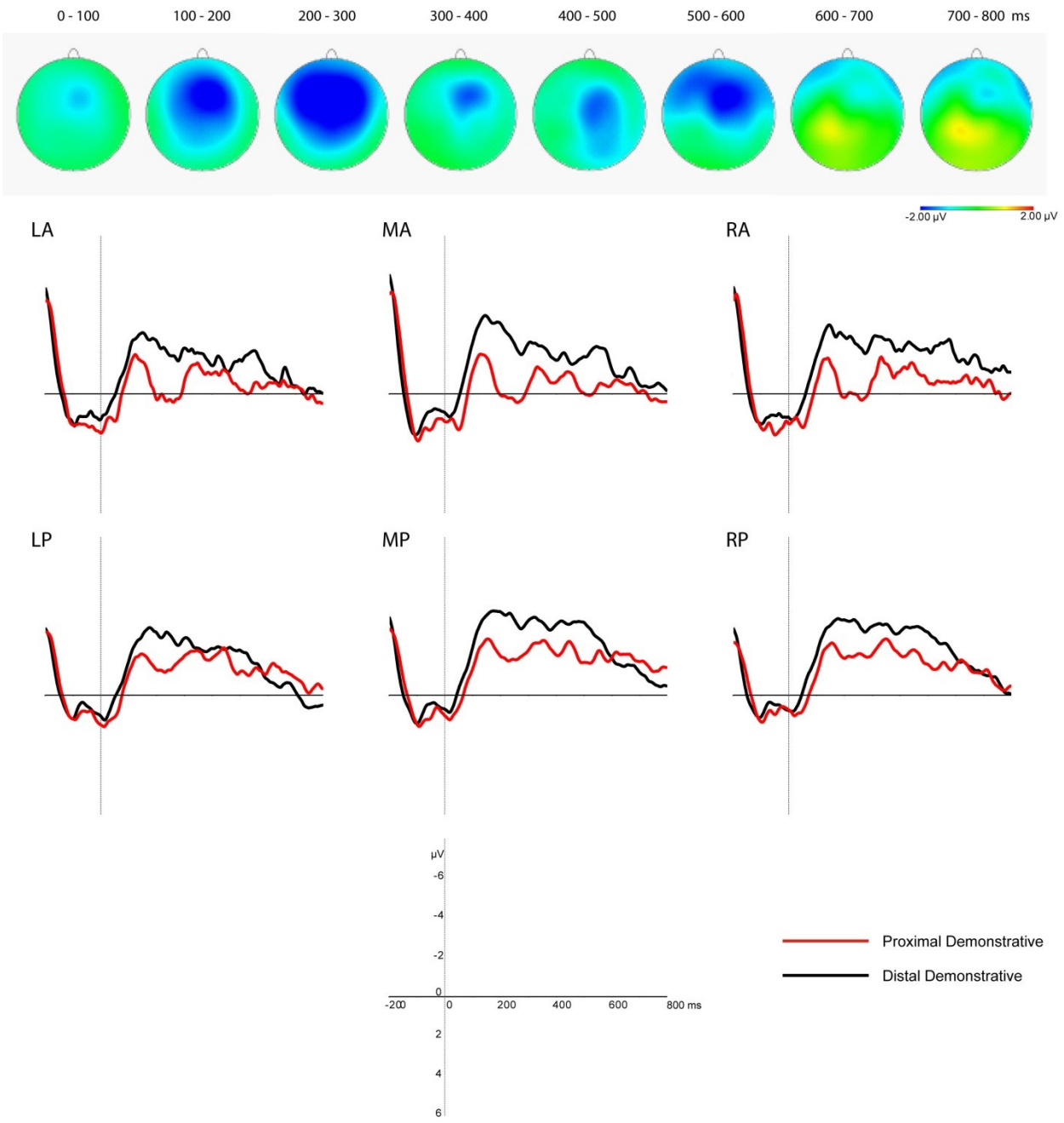

Figure 4. *Grand average waveforms time-locked to the onset of the demonstrative (proximal vs. distal) for the picture stimuli with a sagittal orientation of referents in Experiment 1. The electrode locations are indicated and can be found in Figure 3. The topographic plots show the corresponding voltage differences between the two conditions, for 100 ms time-windows.*

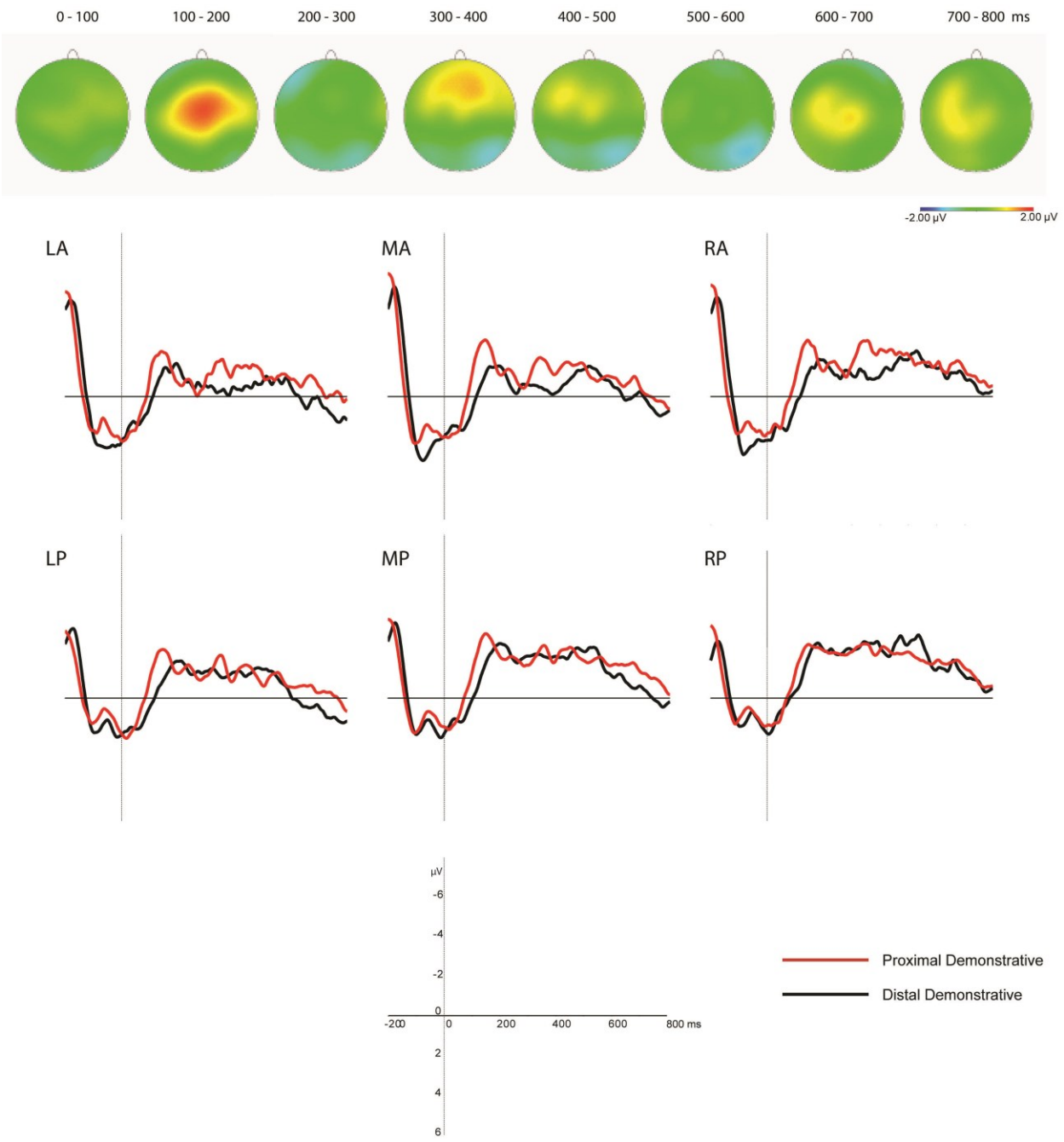

Figure 5. *Grand average waveforms time-locked to the onset of the demonstrative (proximal vs. distal) for the picture stimuli with a lateral orientation of referents in Experiment 1. The topographic plots show the corresponding voltage differences between the two conditions, for 100 ms time-windows.*

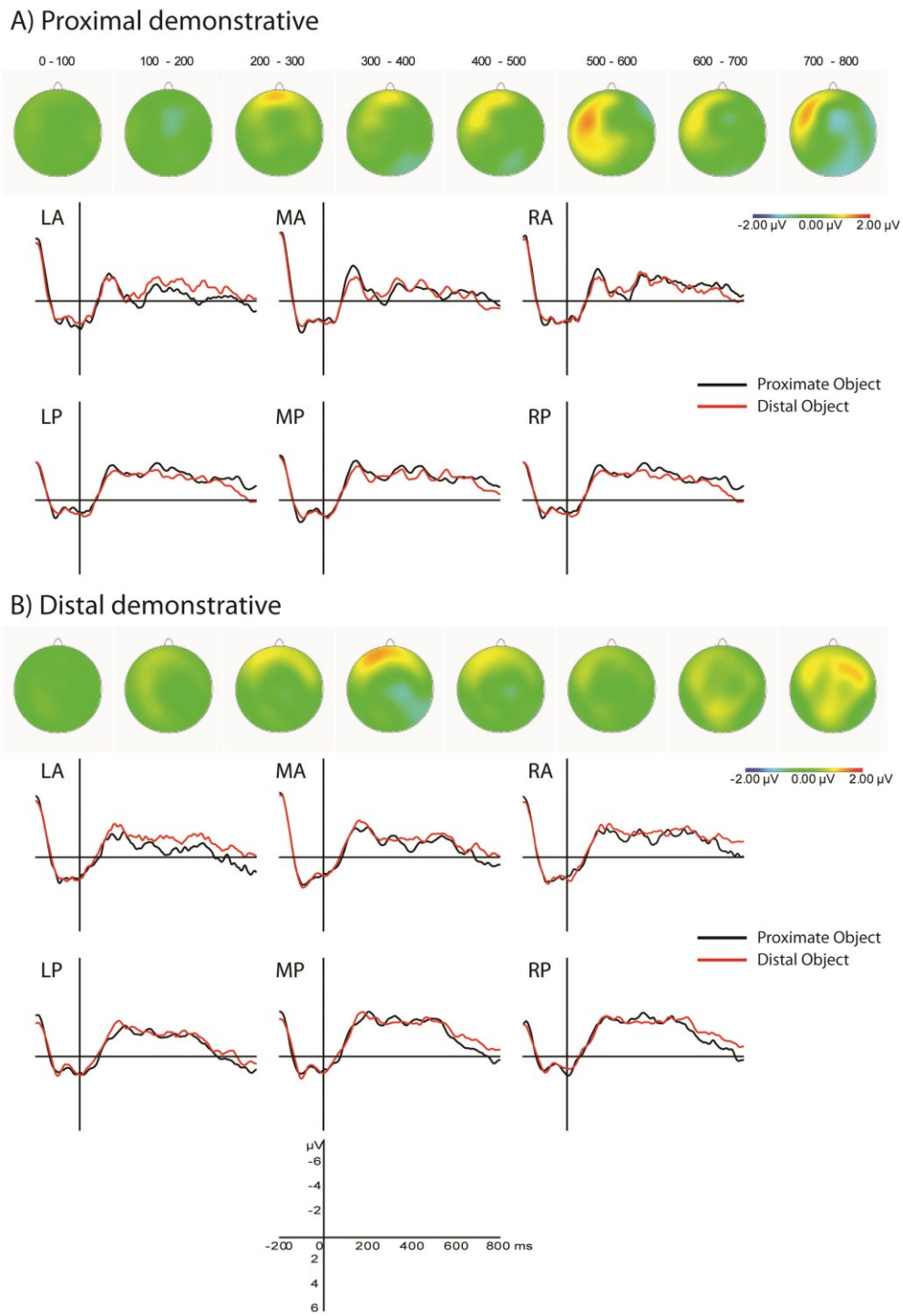

Figure 6. *Grand average waveforms time-locked to the onset of the demonstrative (panel A: proximal demonstrative; panel B: distal demonstrative) for the picture stimuli in which the referent was close to the speaker (black line) or relatively distal (red line), collapsed across Orientation. The topographic plots show the corresponding voltage differences between the two conditions, for 100 ms time-windows.*

Table 3 shows the outcome of the ERP analyses time-locked to the onset of the noun. A large N400 effect was found comparing incongruent to congruent referential nouns, which starts out as an effect that is larger in centro-parietal regions in early time-windows and is larger in anterior regions in late time-windows. Figure 7 shows the grand average waveforms and the topographic plots.

Finally, we directly compared the effect found time-locked to the demonstrative (as presented in Figure 4) to the congruity effect found on the noun (as presented in Figure 7). The dependent variable in this analysis was the difference wave, both for the demonstrative effect and for the noun effect, in the different time-windows and the independent variables used were Word Class (Demonstrative versus Noun) and Quadrant. This analysis yielded only a significant Word Class x Quadrant interaction effect in the 200-300 ms time-window, F (2,39) = 4.43, p = .022. However, follow-up analyses for the separate quadrants did not reveal any significant main effect of Word Class (all p's > .15). Thus, both the ERP effect on the demonstrative and the ERP effect on the noun had a negative directionality, were widespread over the scalp, and took place in similar time-windows. Due to the similar time course and negative directionality, we interpret a more negative wave in the demonstrative manipulation as reflecting a higher processing cost (cf. Stevens & Zhang, 2013).

Discussion

Experiment 1 contrasted and tested two theoretical accounts of demonstrative reference by simultaneously presenting visual and auditory stimuli while recording the EEG of native speakers of Dutch.

The results from the pretest, which tapped into participants' linguistic intuitions on demonstrative use, were in line with the egocentric proximity account in that a proximal demonstrative was used for objects relatively close to the speaker, and a distal demonstrative for objects relatively remote from the speaker's location, irrespective of the spatial orientation of objects. These results also resemble the congruency judgments participants made in the study by Stevens and Zhang (2013). Piwek et al., (2008) termed this intuitive belief on demonstrative use "the folk view" on proximal and distal demonstratives. Interestingly, the current ERP results tell a different story.

The egocentric proximity account (e.g., Anderson & Keenan, 1985; Clark & Sengul, 1978; Diessel, 1999; Fillmore, 1982; Halliday & Hasan, 1977; Lakoff, 1974; Lyons, 1977; Rauh, 1983; Russell, 1940) predicted a Demonstrative x Distance interaction. According to this account, perceiving a proximal demonstrative in reference to an object remote from the speaker, and perceiving a distal demonstrative in reference to an object close to the speaker, would be experienced as caused by a violation, leading to a higher processing cost. The orientation of objects in a lateral or sagittal plane would not influence demonstrative comprehension. The current ERP data reject this account, because there was no Demonstrative x Distance interaction, and the orientation of objects played a crucial role. When objects were oriented in a sagittal plane between speaker and participant, a more negative wave was found for the distal compared to the proximal demonstrative, irrespective of whether the referent was relatively close to or remote from the speaker. On the other hand, when objects were oriented on a lateral plane, no difference was found for the ERPs to the proximal compared to the distal demonstratives. The egocentric proximity account cannot explain these findings.

Table 3. *Results of the ERP analyses time-locked to the onset of the noun in Experiment 1.*

	100-200	200-300	300-400	400-500	500-600	600-700	700-800
EXPERIMENT 1							
Overall							
Noun x Quadrant	< 1	2.09	4.76*	< 1	5.30*	3.47*	1.21
Noun Main Effect	4.14	27.73***	44.87***	18.20***	2.27	1.33	< 1
Follow-up							
Noun Main Effect							
In LA Quadrant	1.96	12.86**	29.01***	22.86***	9.05**	6.33*	2.54
In RA Quadrant	4.17	13.16**	29.04***	15.21**	4.25	1.67	< 1
In LP Quadrant	8.03*	38.50***	45.70***	18.89***	< 1	< 1	3.44
In RP Quadrant	< 1	6.57*	36.29***	6.21*	< 1	< 1	< 1
In Vertical Midline	10.72**	30.93***	41.92***	20.43***	2.57	< 1	< 1

* $p < .05$; ** $p < .01$; *** $p < .001$

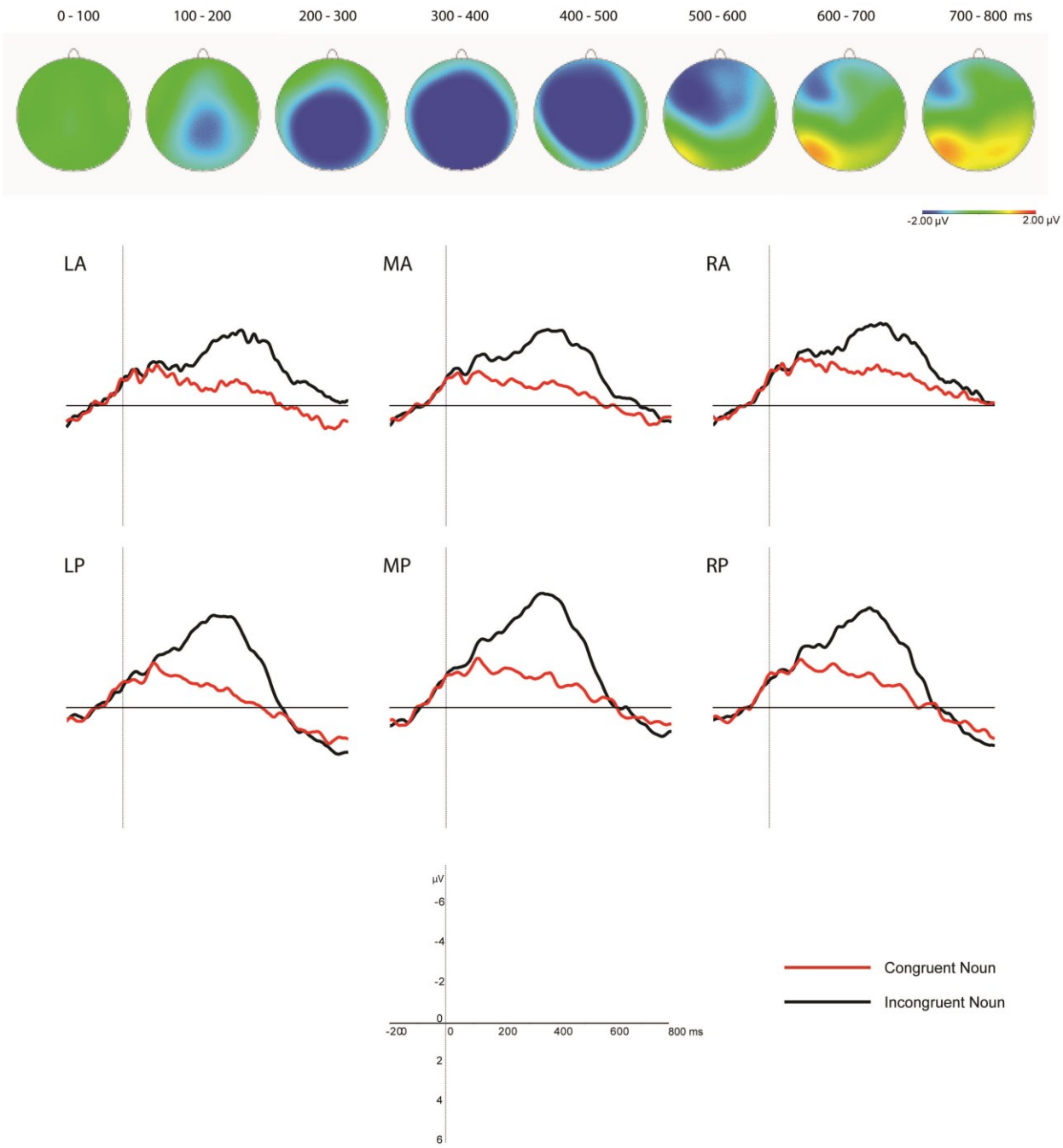

Figure 7. *Grand average waveforms time-locked to the onset of the noun (incongruent vs. congruent) in Experiment 1. The topographic plots show the corresponding voltage differences between the two conditions, for 100 ms time-windows.*

The results of Experiment 1 are more in line with the dyad-oriented account, as put forward by Jungbluth (2003). Indeed, when speaker and participant were face-to-face as in a

conversational dyad, and both potential referents were located in the shared space between them, the ERPs showed a more negative wave for the distal compared to the proximal demonstrative irrespective of the location of the referent. This is exactly what the dyad-oriented account predicted: distal demonstratives are more inappropriate in reference to an object in shared space, and therefore recruit more processing resources. It may seem surprising that no ERP difference was found when objects were oriented laterally, particularly for the proximate object which was located in the space between speaker and participant. However, because the picture stimuli were presented centrally on the screen from the perspective of the participant, the bodily orientations of speaker and implicit addressee were not aligned (see Figure 1), which may be a prerequisite for creating shared space. If this is the case, aligning speaker and participant would lead to a preference for a proximal demonstrative for the object close to the speaker in the lateral orientation of objects. We tested this possibility in Experiment 2, by aligning the speaker in the pictures with the position of the addressee participant.

Furthermore, Coventry et al. (2008) have argued that there may be a parallel between the linguistic encoding of space as evident in the proximal-distal demonstrative distinction and the neural perceptual encoding of space as divided into a peripersonal (within the reach of one's arm) and an extrapersonal (beyond the reach of one's arm) region of space. Proximal demonstratives would be used to refer to objects in peripersonal space, and distal ones for objects in extrapersonal space (Coventry et al., 2008). Although this hypothesis cannot explain why we did not find an ERP difference for demonstratives in the picture stimuli with a lateral orientation of objects, one could argue that, even though in the sagittal pictures one object was relatively more distal from the speaker than the other, they were still both in peripersonal space. This

possibility was also tested in Experiment 2, by enlarging the space between the two objects in the sagitally oriented picture stimuli.

Experiment 2

Method

Participants

Twenty-six participants (mean age 20.3, range 18-25, all female) took part in the experiment. They were all right-handed as assessed by a Dutch translation of the Edinburgh Inventory for hand dominance (Oldfield, 1971). Data from two participants was discarded due to a large number of trials that contained eye blinks and movement artifacts. All participants were Dutch, studying in Nijmegen, and Dutch was their native language. They had normal or corrected-to-normal vision, no language or hearing impairments or history of neurological disease. They provided written informed consent and were paid for participation.

Materials and Pretest of Materials

There were two changes in Experiment 2 compared to Experiment 1. First, the picture stimuli with a lateral orientation of objects were now presented with the speaker central on the screen, thus aligned with the participant, who was sitting centrally in front of the screen (as in Experiment 1). Second, the pictures with a sagittal orientation of objects were modified using Adobe Photoshop such that the distance between the two objects in the pictures became larger and the distal object was clearly out of reach of the speaker (see Figure 8).

The 120 target pictures used in Experiment 1 but adapted in the two ways described above were presented in a pretest that was similar to the pretest described for Experiment 1.

The results of this second pretest (9 subjects, 4 male, mean age 23.9, not participating in the main experiment) were similar to the results of the pretest that preceded Experiment 1. Whenever the speaker pointed to the close object, participants used a proximal demonstrative ("this") in 96.80 % of the cases. When the speaker pointed to the remote object, participants used a distal demonstrative ("that") in 99.17 % of the cases. This preference was again irrespective of the orientation (lateral vs. sagittal) of objects in the pictures.

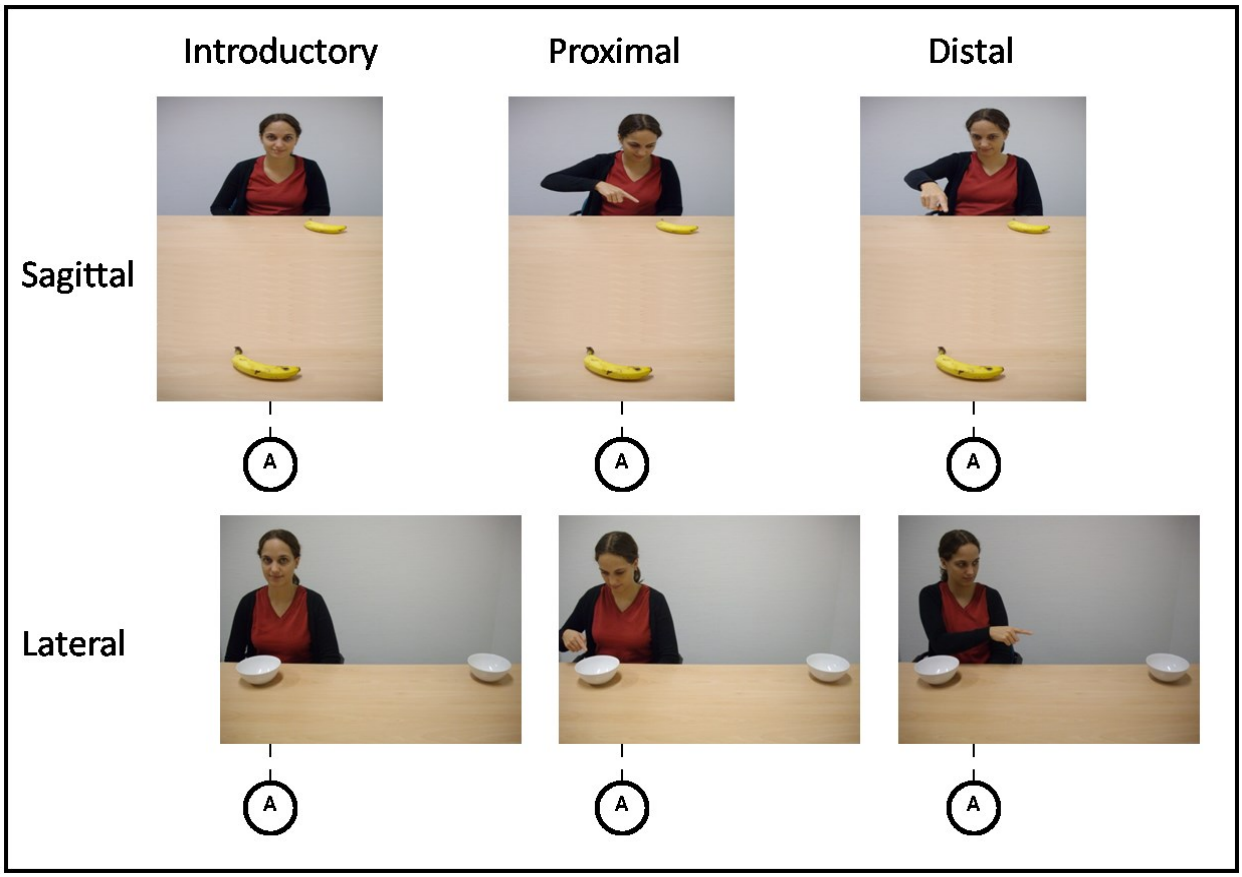

Figure 8. *Subset of stimuli used in Experiment 2. Each row contains an introductory picture, a target picture in which the speaker refers to the proximate object, and a target picture in which she refers to the distal object. The top row shows pictures with a sagittal orientation of objects. The bottom row shows pictures with a lateral orientation of objects. The position of the participant addressee is marked below the pictures.*

Procedure, EEG Recording and analysis

The experimental procedure, and the recording and analysis of the EEG data were identical to the procedure described for Experiment 1. As in Experiment 1, trials containing ocular or muscular artifacts were not taken into consideration in the averaging process (7.83 % of all data). Participants correctly identified 94.1 % of all catch trials.

Results

Table 4 gives an overview of the results of the time-window analysis of the ERPs time-locked to the onset of the demonstrative in Experiment 2. The overall analysis revealed a significant three-way interaction effect between Demonstrative, Orientation, and Quadrant in time-windows between 100 and 300 ms after demonstrative onset. Follow-up analyses revealed a significant interaction effect of Demonstrative x Quadrant in these time-windows for the picture stimuli with a sagittal orientation of referents. This effect started earlier and was larger in anterior than in posterior quadrants. Figure 9 shows this effect, which denotes a more negative wave for the distal demonstrative compared to the proximal demonstrative, having an anterior scalp distribution. No such effect was found for the pictures with a lateral orientation of objects. The follow-up analysis for the picture stimuli with a lateral orientation of objects did yield a significant main effect of Demonstrative in an early and a late time-window. Figure 10 shows the grand average waveforms and the topographic plots corresponding to the demonstrative contrast for the picture stimuli with a lateral orientation of referents. The small three-way interaction with Distance in the overall analysis did not yield any significant follow-up effects related to the Demonstrative factor. Table 4 and Figure 11 show that the Demonstrative x Distance interaction predicted by the egocentric proximity account was again not present in the data.

Table 4. *Results of the ERP analyses time-locked to the onset of the demonstrative in Experiment 2.*

	100-200	200-300	300-400	400-500	500-600	600-700	700-800
EXPERIMENT 2							
Overall							
Demonstrative x Orientation x Quadrant	7.70**	7.13*	3.06	2.01	1.31	< 1	1.50
Demonstrative x Orientation x Distance	< 1	< 1	< 1	6.11*	1.64	2.28	< 1
Demonstrative x Distance	< 1	2.41	1.02	< 1	< 1	< 1	< 1
Demonstrative x Orientation	7.19*	7.74*	6.70*	4.04	2.54	1.46	< 1
Demonstrative x Quadrant	1.04	2.49	2.41	1.27	< 1	< 1	1.62
Demonstrative	< 1	6.84*	< 1	< 1	< 1	5.39*	1.17
Lateral Orientation							
Dem x Quadrant	2.57	1.46	3.09	3.10	< 1	< 1	< 1
Dem x Distance	< 1	< 1	< 1	3.40	1.20	< 1	< 1
Demonstrative	2.71	< 1	4.94*	2.58	3.91	4.84*	1.70
Sagittal Orientation							
Dem x Quadrant	7.32**	7.09*	< 1	< 1	1.47	1.49	2.68
Dem x Distance	< 1	3.11	1.76	3.27	< 1	4.40*	< 1
Demonstrative	2.99	12.07**	3.26	1.14	< 1	< 1	< 1
In LA Quadrant	9.00**	22.30***	1.98	< 1	1.07	< 1	< 1
In RA Quadrant	4.92*	12.65**	3.19	1.37	1.40	< 1	< 1
In LP Quadrant	< 1	6.16*	2.13	< 1	< 1	< 1	< 1
In RP Quadrant	< 1	3.67	3.84	< 1	< 1	1.05	1.28
In Vertical Midline	3.59	12.02**	3.78	1.38	< 1	< 1	< 1

* $p < .05$; ** $p < .01$; *** $p < .001$

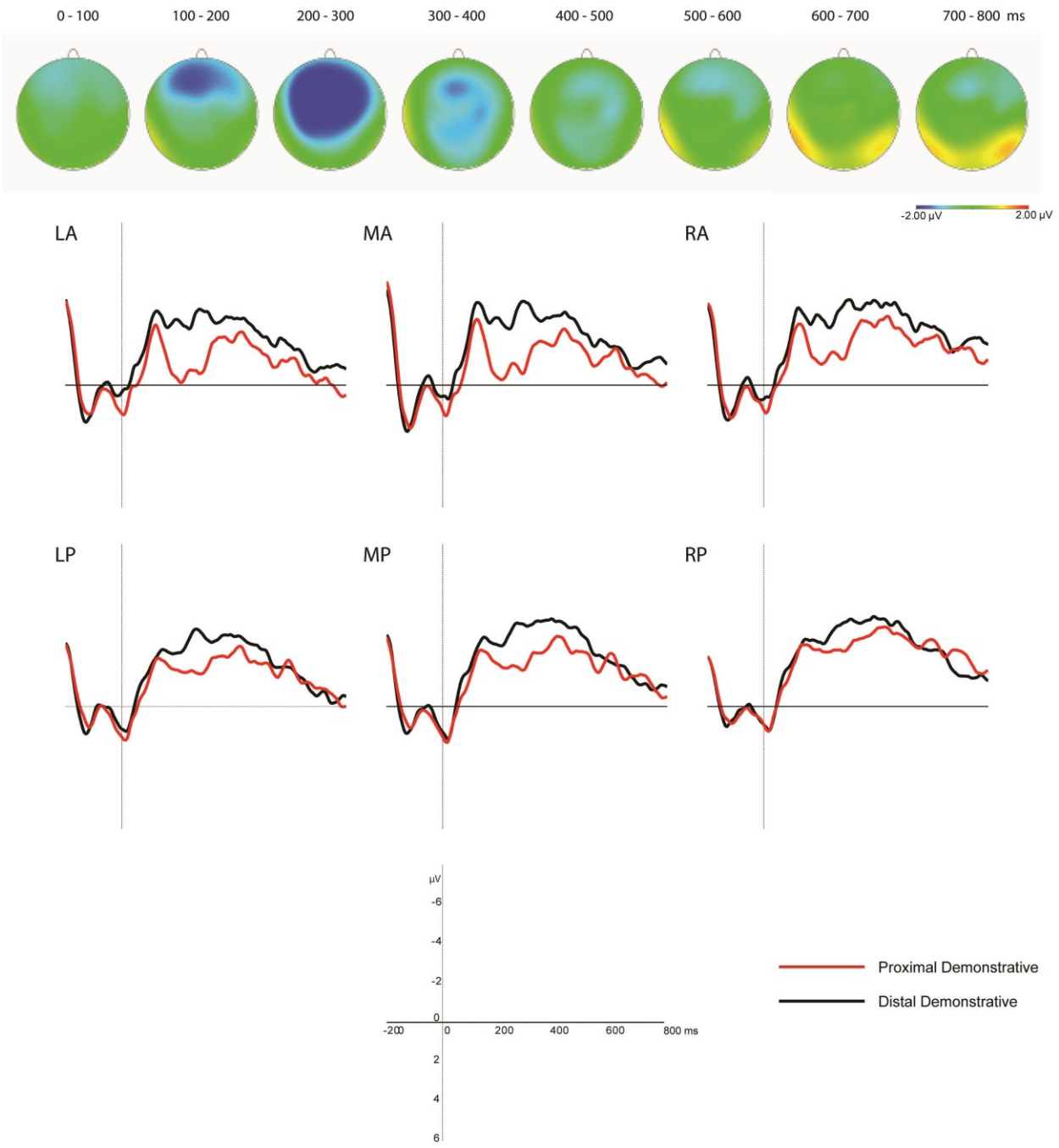

Figure 9. *Grand average waveforms time-locked to the onset of the demonstrative (proximal vs. distal) for the picture stimuli with a sagittal orientation of referents in Experiment 2. The topographic plots show the corresponding voltage differences between the two conditions, for 100 ms time-windows.*

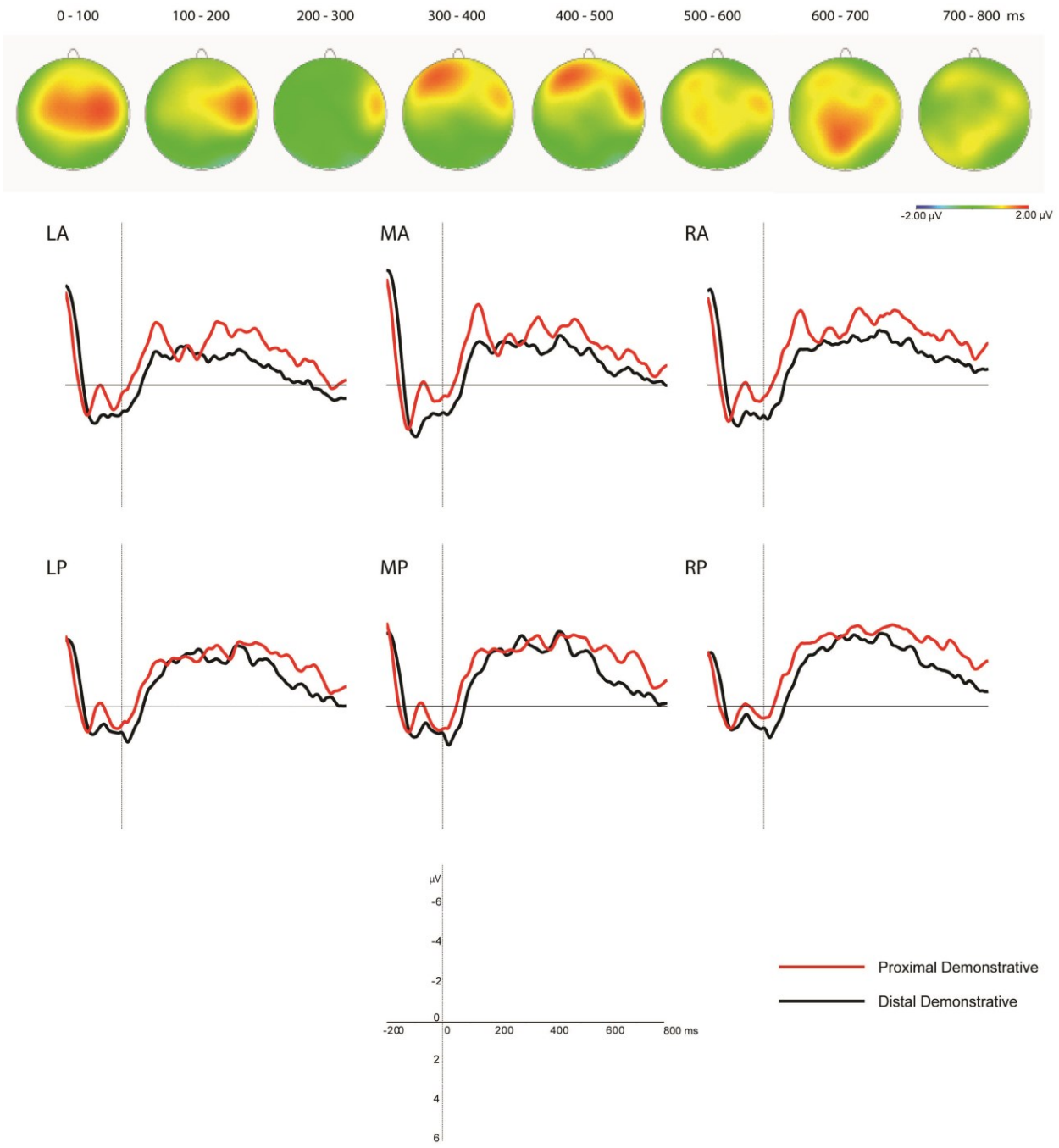

Figure 10. *Grand average waveforms time-locked to the onset of the demonstrative (proximal vs. distal) for the picture stimuli with a lateral orientation of referents in Experiment 2. The topographic plots show the corresponding voltage differences between the two conditions, for 100 ms time-windows.*

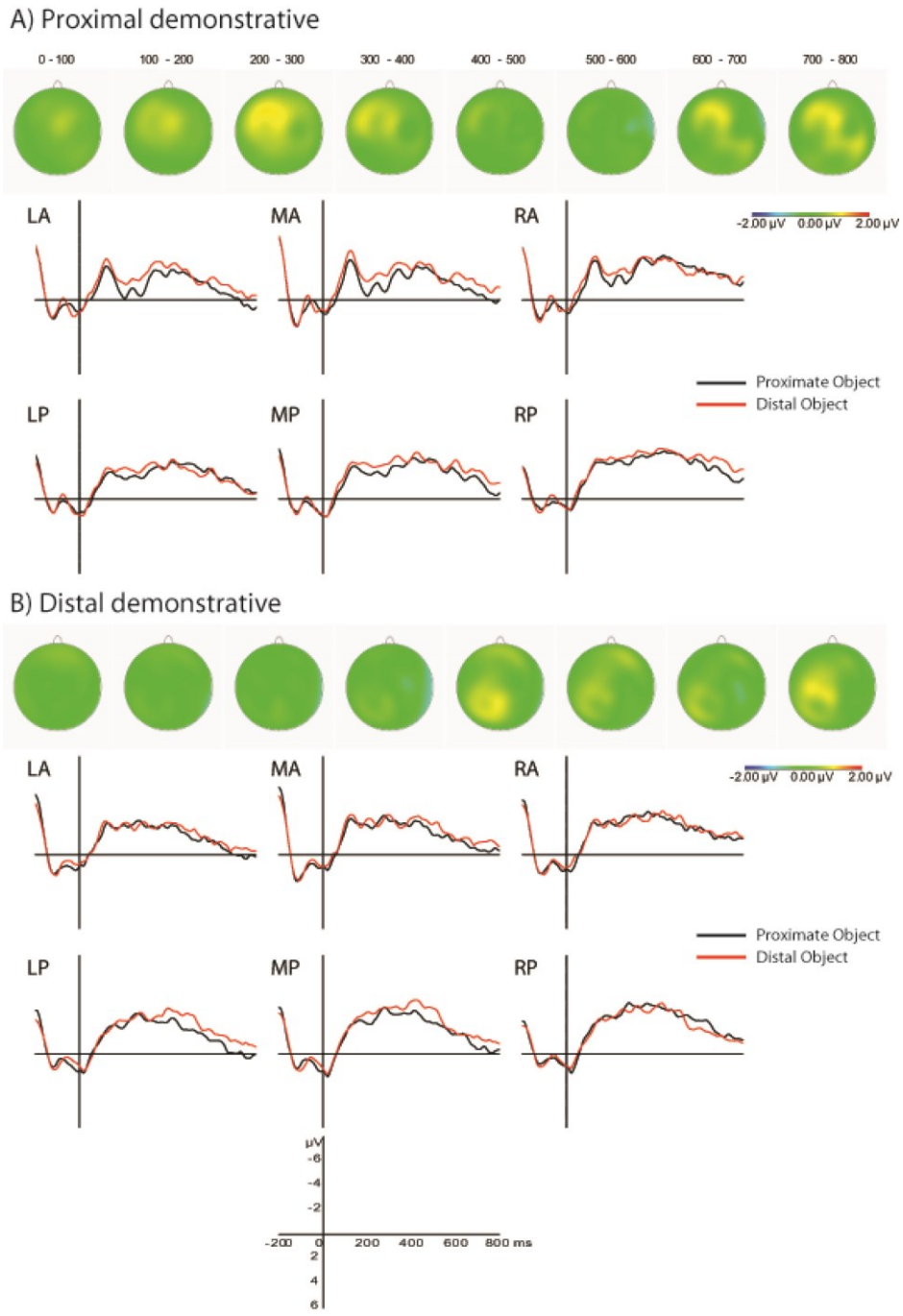

Figure 11. *Grand average waveforms time-locked to the onset of the demonstrative (panel A: proximal demonstrative; panel B: distal demonstrative) for the picture stimuli in which the referent was close to the speaker (black line) or relatively distal (red line), collapsed across Orientation. The topographic plots show the corresponding voltage differences between the two conditions, for 100 ms time-windows.*

Table 5. *Results of the ERP analyses time-locked to the onset of the noun in Experiment 2.*

	100-200	200-300	300-400	400-500	500-600	600-700	700-800
EXPERIMENT 2							
Overall							
Noun x Distance x Quadrant	2.81	5.44*	2.54	2.47	1.93	< 1	1.23
Noun x Quadrant	< 1	< 1	< 1	3.78*	20.56***	23.17***	18.54***
Noun x Orientation	3.50	9.78**	8.60**	1.91	< 1	1.23	1.51
Noun x Distance	< 1	< 1	< 1	< 1	1.79	4.82*	1.92
Noun Main Effect	15.55**	38.37***	62.96***	61.98***	21.84***	6.49*	< 1
Follow-up							
Noun Main Effect							
In LA Quadrant	5.03*	19.72***	43.57***	58.14***	65.77***	45.54***	6.56*
In RA Quadrant	6.94*	21.79***	42.13***	54.99***	48.37***	25.27***	5.91*
In LP Quadrant	12.48**	35.62***	55.45***	35.65***	1.12	< 1	3.64
In RP Quadrant	18.93***	45.24***	64.50***	42.30***	1.64	< 1	1.85
In Vertical Midline	14.99**	37.32***	54.41***	51.41***	19.05***	6.54*	< 1
For Proximate Objects	3.66	18.35***	29.93***	42.58***	20.05***	9.27**	1.68
For Distal Objects	8.83**	19.42***	60.71***	38.77***	6.69*	< 1	< 1
In Lateral Orientation	1.29	8.70**	13.93**	15.20**	4.13	< 1	< 1
In Sagittal Orientation	14.54**	38.34***	65.02***	41.64***	17.77***	6.18*	1.54

* $p < .05$; ** $p < .01$; *** $p < .001$

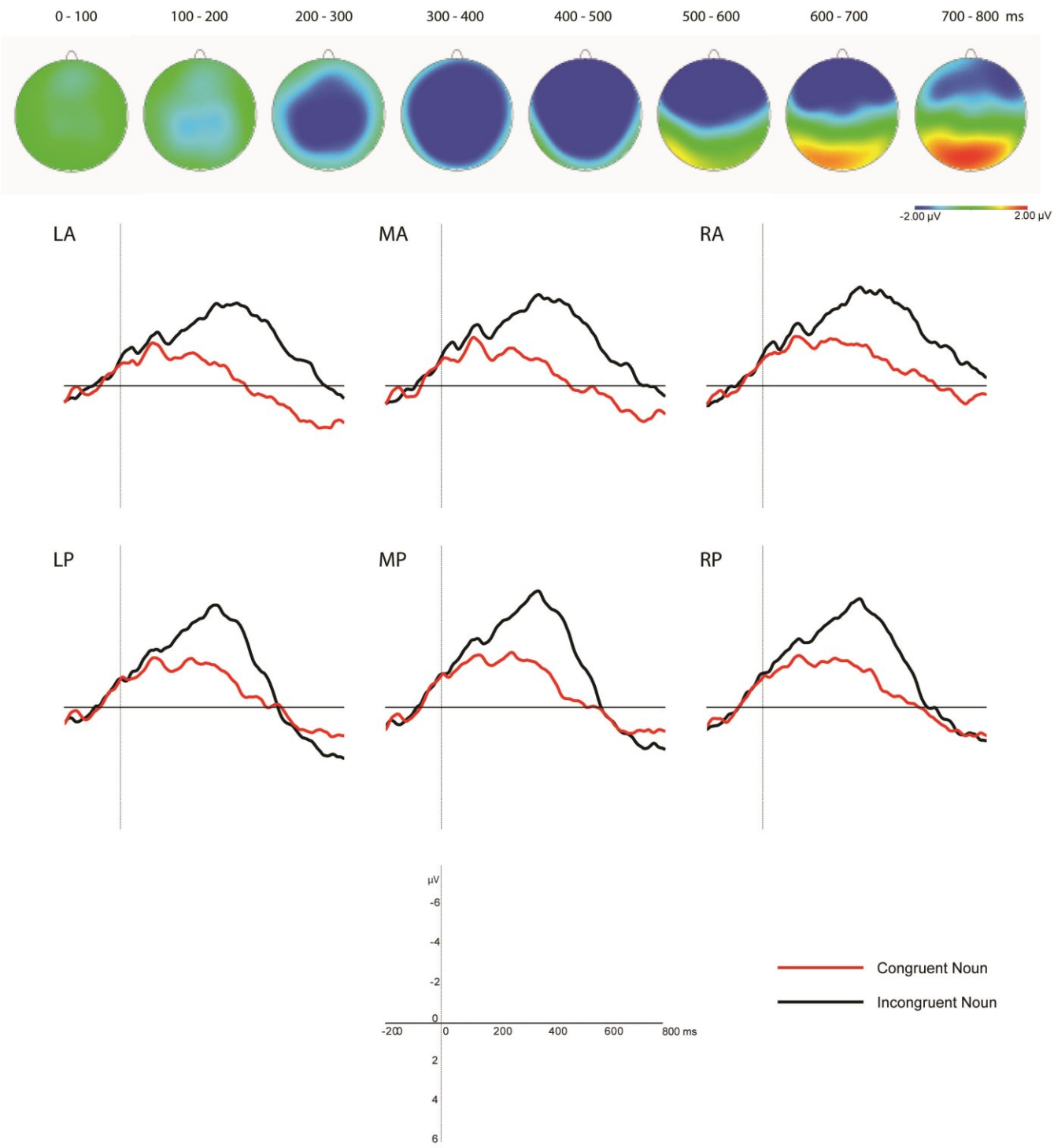

Figure 12. *Grand average waveforms time-locked to the onset of the noun (incongruent vs. congruent) in Experiment 2. The topographic plots show the corresponding voltage differences between the two conditions, for 100 ms time-windows.*

Table 5 shows the outcome of the ERP analyses time-locked to the onset of the noun in Experiment 2. Similar to Experiment 1, a large N400 effect was found when comparing incongruent to congruent referential nouns, which started out as an effect that was larger in centro-parietal regions in early time-windows and was larger in anterior regions in late time-windows. The higher-order interaction effects denote that the effect of Noun was larger posteriorly in early time-windows and larger anteriorly in later time-windows, that it started slightly later when the referent was proximate compared to when it was distal to the speaker, and that it started slightly earlier and lasted longer when the referents were in sagittal orientation compared to when they were laterally oriented. Figure 12 shows the overall grand average waveforms and the topographic plots.

Identical to Experiment 1, we directly compared the effect found time-locked to the demonstrative (as presented in Figure 9) to the congruity effect found on the noun (as presented in Figure 12). This analysis yielded significant Word Class x Quadrant interaction effects in the 200-300 ms time-window, F (2,50) = 3.96, p = .022, the 500-600 ms time-window, F (2,44) = 5.45, p = .008, and the 600-700 ms time-window, F (2,44) = 4.76, p = .015. The early effect (200-300 ms) reflected trends towards a significant main effect of Word Class in both the left and right posterior quadrants (p = .054 and .062 respectively) but not in the other quadrants. The late effects reflected significant main effects of Word Class in the left anterior quadrant (p = .001 and .003 in the two subsequent late time-windows), the right anterior quadrant (p = .011 and .010 respectively) and over the vertical midline channels (p = .033 and .036 respectively). These findings reflect that the demonstrative effect was slightly larger in anterior regions in an early time-window but smaller in later time-windows compared to the congruity effect on the noun.

Discussion

Experiment 2 further tested the dyad-oriented account of demonstrative reference. In Experiment 1, it was found that participants did not prefer one demonstrative over another when a speaker referred to one of two objects that were oriented laterally in front of her, as indicated by an absence of a significant ERP difference across the auditory presentation of different demonstratives. This raised the question whether the bodily orientations of speaker and (implicit) addressee need to be aligned in order to create a notion of shared space. An alternative possibility was that the dyad-oriented account only holds for situations in which all possible referents are located within the shared dyadic space in between speaker and addressee. The findings of Experiment 2 were in line with this latter possibility. When the bodily orientations of speaker and participant were aligned, and the speaker referred to the object close to her while the other object was remote in a lateral plane, no proximal demonstrative preference was detected. If anything, the small negative-going main effects suggest a preference for a distal demonstrative for both objects in the lateral plane. Thus, these findings further specify the dyad-oriented account in underlining that people do indeed prefer a proximal demonstrative for referents in the shared space between speaker and addressee, but only if all possible referents are located within that space. In other words, the presence of another possible referent outside of the dyad, which introduces a lateral axis and as such a possible boundary between speaker and addressee, eliminates their experience of sharing the extra-linguistic space in between them, by creating a different division of that space. When space is no longer shared, addressees may prefer a distal demonstrative for all potential referents.

General Discussion

In two ERP experiments, participants watched images of a speaker referring to one of two objects while they listened to her referential speech. In order to contrast and test two theoretical accounts on demonstrative reference, participants' concurrently recorded electroencephalograms were analyzed to

find out whether they had a preference for one demonstrative over another as a function of the proximity of the object-referent to the speaker and the orientation of objects in space. It was found that participants preferred a proximal demonstrative to all objects located in the shared space between speaker and addressee, but only when there was no other possible referent outside of this space. Whenever there was a potential referent outside of the shared space, participants either had no preference (as in Experiment 1) or preferred a distal demonstrative (as in Experiment 2) both for referents close to the speaker as well as for referents distal from the speaker. The theoretical importance of these findings will now be discussed in the light of the theoretical views on demonstrative reference presented in the introduction.

The first theoretical view on demonstratives put forward was termed the egocentric proximity account. This view (e.g., Anderson & Keenan, 1985; Clark & Sengul, 1978; Diessel, 1999; Fillmore, 1982; Halliday & Hasan, 1977; Lakoff, 1974; Lyons, 1977; Rauh, 1983; Russell, 1940) predicted that addressees would prefer a proximal demonstrative for referents physically close to the speaker and a distal demonstrative for referents physically remote from the speaker. Our ERP data strongly falsify this view on demonstrative reference and cannot be interpreted in line with it in any way, because the data show that participants simply did not base their online demonstrative comprehension on the referent's relative proximity to the speaker. In line with the egocentric proximity account, Coventry et al. (2008) proposed that speakers would prefer a proximal demonstrative for referents in peripersonal space and a distal demonstrative for referents in extrapersonal space. Experiment 2 falsifies this proposal, in showing that participants may prefer a proximal demonstrative for referents in the speaker's extrapersonal space. There are other theoretical and empirical arguments against the suggested parallel between perceptual and linguistic encoding of space (see e.g., Bonfiglioli et al., 2009; Kemmerer, 1999; Kirsner, 1993). As argued by Kemmerer (1999, p. 47), for instance, the mere existence of many languages in which there are three or more types of demonstrative shows that demonstrative systems need not correspond to the near-far perceptual contrast.

Nevertheless, the egocentric proximity view was in line with native speakers' linguistic intuitions on demonstratives, as evident in the pre-tests of the current study and in linguistic judgments made by participants in previous studies (e.g., Stevens & Zhang, 2013). Assuming that reference grammars often partly or fully rely on linguistic intuitions, this may explain why the egocentric proximity account is omnipresent in such grammars (Diessel, 2005). However, observational (Enfield, 2003; Hanks, 1990; Jungbluth, 2003; Piwek et al., 2008) and online processing data (this study) strongly suggest that such linguistic intuitions are not reliable (see Clark & Bangerter, 2004, for a similar argument). Although there may be instances in which speakers use demonstratives on the basis of the physical distance of referents, possibly when speaker and addressee sit side-by-side (Jungbluth, 2003), this does not imply that they are egocentric in only considering the distance of referents from their own physical location without taking their addressee's location and perspective into account. Below we will argue in favor of a sociocentric view on reference to replace this egocentric view (cf. Clark & Bangerter, 2004; Jungbluth, 2003; Weinrich, 1988).

The current ERP results are not in line with a previous EEG study on the comprehension of demonstratives either. Stevens and Zhang (2013) only found an ERP effect for situations in which a speaker referred to an object close to the addressee and speaker and addressee shared gaze at this object. They interpreted this effect as denoting a preference for a distal demonstrative in this situation, in line with participants' linguistic intuitions. In our study we found a preference for a proximal demonstrative for the referent close to the addressee. One possible explanation for this discrepancy across studies may be that overhearing referential speech, as in Stevens and Zhang (2013), may require different processing mechanisms compared to being the (implicit) addressee of referential speech and gesture (see e.g., Schober & Clark, 1989) as in the current study. An alternative explanation for the finding in Stevens and Zhang (2013), and the absence of ERP effects for any other contrast in their study, may be the fact that their participants on every trial made congruency judgments on the relation between the auditorily

presented demonstrative and the visually presented image. Participants' use of linguistic intuitions and the fact that they were not naïve to the goal of the experiment may have prevented a natural processing of the presented demonstratives.

Our data are most in line with the dyad-oriented account of demonstrative reference (Jungbluth, 2003; see also Weinrich, 1988, and Laury, 1996). For the stimuli with a sagittal orientation of objects, the dyad-oriented account exactly predicted the pattern of results we found. Indeed, irrespective of the proximity of the referent to the speaker, the ERP results indicate that addressees preferred a proximal demonstrative for the referents within the dyad. However, the dyad-oriented account cannot explain why participants did not prefer a proximal demonstrative for the object close to the speaker in the lateral orientation of objects, in both experiments. This object was located in the space in between speaker and participant, and did not yield a preference for a proximal demonstrative, even when the bodily orientations of speaker and addressee were aligned as in Experiment 2. We here propose a shared-space account of demonstrative reference, which can explain these findings.

The shared-space account underlines that in human interaction physical space is transformed into meaningful space (Enfield, 2003, p. 88; Hanks, 1990; Kendon, 1977; 1990; 1992; Scheflen & Ashcraft, 1976). In the case of demonstrative reference, this means that at the time of a certain referential utterance during a conversation, interlocutors may experience some part of the extra-linguistic space as being shared (somewhat similar to Enfield's *here-space*), which would elicit the use of proximal demonstratives for referents within the shared space. Being in a face-to-face situation with an alignment of bodily orientations, as in a conversational dyad, the space within the dyad is likely to be experienced as shared (Jungbluth, 2003). However, the presence of another possible referent in the same physical (lateral) axis as the actual referent may create a boundary that separates speaker and addressee and as such eliminates the experience of sharing space, consequently leading to a preference for a distal demonstrative or to no particular preference at all. This lateral barrier may have been experienced as

more salient in Experiment 2 compared to Experiment 1, because it was orthogonal with the sagittal axis in the former, therefore leading to a stronger preference for a distal demonstrative for objects located on the lateral axis in Experiment 2. Physical boundaries may also play a role in eliminating the construal of shared space. In retrospect, this explains why participants in a building task reported by Clark and Krych (2004) used significantly more distal than proximal demonstratives in referring to their building blocks when there was a physical barrier in between of the two partners. Indeed, this preference decreased drastically when the physical barrier was removed. The shared-space account does not claim that the space interlocutors share can uniquely be in between of their aligned bodies. Instead, we propose that interlocutors in a conversation build up shared space throughout the course of a conversation, taking into account previous interactions and mutually shared knowledge.

The shared-space account moves away from egocentric explanations of deixis and pleads for a sociocentric approach instead (cf. Clark & Bangerter, 2004; Clark & Wilkes-Gibbs, 1986; Jungbluth, 2003; Weinrich, 1988). It underlines that referring is a collaborative and cooperative process (Clark, 1996; Clark & Bangerter, 2004; Clark & Wilkes-Gibbs, 1986) in which speakers are not blind to their addressees. Indeed, speakers actively take into account the location and bodily orientation of their addressee in interaction (Kendon, 1990). Kendon (1990; 1992) underlines that interlocutors often use their bodies to separate their shared space of engagement from the outside world, creating a so-called joint transactional segment. Özyürek (2000; 2002) showed that speakers also take into account the location and orientation in space of their addressees, and as such the shared space between speaker and addressee, when designing their co-speech representational gestures. Liddell (1995; 2000) showed that signers of American Sign Language take into account the location of their addressee in determining the orientation and direction of pronouns and indicating verb signs. Here we show that also in demonstrative reference the bodily orientation of speaker and addressee plays an important role and that mutual awareness of this is likely. As such, we follow Enfield (2003, p. 115) in arguing that all demonstratives

are both speaker-anchored and addressee-anchored. This also fits well with a large body of work indicating that in general, speakers take into account characteristics of their specific recipient or the specific communicative situation when designing their communicatively intended message as transmitted via speech (e.g., Clark, 1996; Brennan & Hanna, 2009), co-speech iconic gestures (e.g., Alibali, Heath, & Myers, 2001), and/or pointing gestures (e.g., Chapter 4 of this thesis).

Demonstrative systems are often said to contain or consist of proximal and distal demonstratives (e.g., Diessel, 2005). However, if physical proximity or distance is not the driving force behind a speaker's choice for a particular demonstrative, this terminology becomes confusing. Nevertheless, the terms *proximal* and *distal* hold when interpreted as referring to psychologically proximal and distal objects, instead of referring to physically proximal and distal objects. Referents within the shared space would then be psychologically proximal and referents outside of the shared space psychologically distal. Demonstratives may indeed fit well within a construal-level theory of psychological distance (Bar-Anan, Liberman, Trope, & Algom, 2007; Liberman & Trope, 2008; Trope & Liberman, 2010) provided that such a theory adopts a sociocentric approach to replace its egocentric starting point.

The current study investigated demonstrative determiners in adnominal use. Future studies are needed to investigate whether the shared-space account also holds for demonstrative adverbs (*here* and *there* in English), in cases where demonstratives are used as pronouns and in different functions (see e.g., Kirsner, 1993; Levinson, 2004), and whether it plays a role in the choice of referring expression from the continuum between definite descriptions and zero anaphora or dropped pronouns.

Finally, the present study confirmed a previous finding by Friedrich and Friederici (2010), who showed that hearing an incorrect label to a visually presented object elicits an N400 effect. Here we extend this finding to triadic referential situations in which speakers may correctly or incorrectly refer to objects in the extra-linguistic space surrounding them. The present study allowed for a descriptive comparison of the timing, directionality and topography of ERP effects on the demonstrative and on the

noun to the same visual materials. Demonstrative effects were found to take place in similar time-windows as the 'canonical' N400 effects on the noun, although the latter effects were larger and lasted longer. Arguably this is due to the demonstrative violations being more subtle and more reversible in that listeners can still relatively easily derive speaker meaning independently of the type of demonstrative used. In addition, both the effect on the noun and the effect on the demonstrative had a similar negative directionality, which led us to interpret a more negative wave as reflecting a higher processing cost (cf. Stevens & Zhang, 2013). Finally, the effects of the demonstrative had a slightly more anterior topographical distribution in an early time-window, which is not uncommon for N400 effects to picture stimuli (Kutas & Federmeier, 2011), compared to the widespread N400 effects of the noun. Alternatively, differences in word class (demonstrative determiner versus noun) may underlie this finding. A comparison of the underlying neural circuitry for demonstratives and co-referential nouns by using imaging techniques with a higher spatial resolution than EEG is an exciting avenue for future research.

To conclude, our study contrasted and tested two theoretical views on demonstrative reference. A dyad-oriented account was most but not perfectly in line with our data. We therefore proposed a shared-space account, which embodies a sociocentric approach to deixis and underlines that, in demonstrative reference, the psychological proximity of a referent may be more important than its physical proximity.

References

Alibali, M. W., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, *44*, 169-188.

Anderson, S. R., & Keenan, E. L., (1985). Deixis. In T. Shopen (Ed.), *Language typology and syntactic description* (pp. 259-308). Cambridge: Cambridge University Press.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother infant and peer-infant interaction. *Child Development*, 55, 1278-1289.

Bar-Anan, Y., Liberman, N., Trope, Y., & Algom, D. (2007). Automatic processing of psychological distance: evidence from a Stroop task. *Journal of Experimental Psychology: General*, *136*, 610-622.

Bonfiglioli, C., Finocchiaro, C., Gesierich, B., Rositani, F., & Vescovi, M. (2009). A kinematic approach to the conceptual representations of *this* and *that*. *Cognition*, *111*, 270-274.

Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science*, *1*(2), 274-291.

Bühler, K. (1934). *Sprachtheorie*. Jena: Fischer.

Burenhult, N. (2003). Attention, accessibility, and the addressee: The case of the Jahai demonstrative ton. *Pragmatics, 13*, 363-379.

Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet.* Mahwah, NJ: Lawrence Erlbaum.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 255, Vol. 63,* 1-174.

Clark, E. V., & Sengul, C. J. (1978). Strategies in the acquisition of deixis. *Journal of Child Language*, *5*(3), 457-475.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet* (pp. 243-268). Hillsdale, NJ: Erlbaum.

Clark, H. H., & Bangerter, A. (2004). Changing ideas about reference. In I. A. Noveck, & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 25-49). Basingstoke: Palgrave Macmillan.

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*, 62-81.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*, 1-39.

Coventry, K. R., Valdés, B., Castillo, A., & Guijarro-Fuentes, P. (2008). Language within your reach: Near–far perceptual space and spatial demonstratives. *Cognition*, *108*, 889-895.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141-168.

Da Milano, F. (2007). Demonstratives in the languages of Europe. In P. Ramat, & E. Roma (Eds.), *Europe and the Mediterranean as Linguistic Areas* (pp. 25-48). Amsterdam: John Benjamins.

Diessel, H. (1999). *Demonstratives. Form, Function, and Grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2005). Distance contrasts in demonstratives. In M. Haspelmat, M. S. Dryer, D. Gil, & B. Comrie (Eds.), *The World Atlas of Language Structures* (pp. 170-173). Oxford: Oxford University Press.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive*

Linguistics, *17*(4), 463–489.

Enfield, N. J. (2003). Demonstratives in space and interaction: Data from Lao speakers and implications for semantic analysis. *Language*, 82-117.

Fillmore, C. J. (1982). Towards a descriptive framework for spatial deixis. In R. J. Jarvella, & W. Klein (Eds.), *Speech, place, & action. Studies in deixis and related topics* (pp. 31-59). Chichester: John Wiley & Sons Ltd.

Friedrich, M., & Friederici, A. D. (2004). N400-like semantic incongruity effect in 19-month-olds: processing known words in picture contexts. *Journal of Cognitive Neuroscience*, *16*, 1465-1477.

Halliday, M. A. K., & Hasan, R. (1977). *Cohesion in English*. London, UK: Longman Group Ltd.

Hanks, W. F. (1990). *Referential practice: Language and lived space among the Maya*. Chicago: University of Chicago Press.

Hanks, W. F. (2005). Explorations in the Deictic Field. *Current Anthropology*, *46*(2), 191-220.

Himmelmann, N. (1996). Demonstratives in narrative discourse: A taxonomy of universal uses. In B. Fox (Ed.), *Studies in anaphora* (pp. 205-254). Amsterdam: John Benjamins.

Hottenroth, P.-M. (1982). The system of local deixis in Spanish. In J. Weissenborn, & W. Klein (Eds.), *Here and there: Cross-linguistic studies on deixis and demonstration* (pp. 133-153). Amsterdam: John Benjamins.

Jakobson, R. (1971). *Selected writings, vol. 2*. The Hague: Mouton.

Jespersen, O. (1922). *Language. Its nature, development, and origin*. London: George Allen & Unwin Ltd.

Jungbluth, K. (2003). Deictics in the conversational dyad: Findings in Spanish and some cross-linguistic outlines. In F. Lenz (Ed.), *Deictic conceptualisation of space, time and person* (pp. 13-40). Amsterdam: John Benjamins.

Kemmerer, D. (1999). “Near” and “far” in language and perception. *Cognition*, *73*, 35-63.

Kendon, A. (1977). Spatial organization in social encounters: The F-formation system. In A. Kendon (Ed.), *Studies in the behavior of social interaction* (pp. 179-208). Lisse: Peter de Ridder Press.

Kendon, A. (1990). *Conducting interaction. Patterns of behavior in focused encounters.* Cambridge: Cambridge University Press.

Kendon, A. (1992). The negotiation of context in face-to-face interaction. In A. Duranti, & C. Goodwin, *Rethinking context: Language as an interactive phenomenon* (pp. 323-334). Cambridge: Cambridge University Press.

Kirsner, R. S. (1993). From meaning to message in two theories: Cognitive and Saussurean views of the Modern Dutch demonstratives. In R. A. Geiger, & B. Rudzka-Ostyn, *Conceptualizations and mental processing in language* (pp. 81-114). Berlin: Mouton de Gruyter.

Küntay, A., & Özyürek, A. (2006). Learning to use demonstratives in conversation: what do language specific strategies in Turkish reveal? *Journal of Child Language*, *33*, 303-320.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, *207*, 203-205.

Lakoff, R. (1974). Remarks on this and that. In M. W. La Galy, R. A. Fox, & A. Bruck (Eds.), *Papers from the tenth regional meeting: Chicago Linguistic Society* (pp. 345-356). Chicago.

Laury, R. (1996). Conversational use and basic meaning of Finnish demonstratives. In A. E. Goldberg (Ed.), *Conceptual structure, discourse and language* (pp. 303-319). Stanford: CSLI publications.

Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.

Levinson, S. C. (2004). Deixis. In L. Horn (Ed.), *The handbook of pragmatics* (pp. 97-121). Oxford: Blackwell.

Liberman, N., & Trope, Y. (2008). The psychology of transcending the here and now. *Science*, *322*, 1201-1205.

Liddell, S. (1995). Real, surrogate and token space: Grammatical consequences in ASL. In K. Emmorey, & J. Reilly (Eds.), *Language, gesture, and space* (pp. 19-43). Hillsdale: Erlbaum.

Liddell, S. (2000). Blended spaces and deixis in sign language discourse. In D. McNeill (Ed.), *Language and gesture* (pp. 331-357). Cambridge: Cambridge University Press.

Lyons, J. (1977). *Semantics. Volume 2*. Cambridge: Cambridge University Press.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*, 97-113.

Özyürek, A. (2000). The influence of addressee location on spatial language and representational gestures of direction. In D. McNeill (Ed.), *Language and gesture* (pp. 64-83). Cambridge: Cambridge University Press.

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, *46*, 688-704.

Peirce, C. S. (1931). *The collected writings of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Piwek, P., Beun, R. J., & Cremers, A. (2008). 'Proximal' and 'distal' in language and cognition: Evidence from deictic demonstratives in Dutch. *Journal of Pragmatics*, *40*, 694-718.

Rauh, G. (1983). Aspects of deixis. In G. Rauh (Ed.), *Essays on deixis* (pp. 9-60). Tübingen: Narr.

Rommetveit, R. (1968). *Words, meanings, and messages: Theory and experiments in psycholinguistics*. New York: Academic Press.

Russell, B. (1940). *An inquiry into meaning and truth*. London: George Allen & Unwin Ltd.

Scheflen, A. E., & Ashcraft, N. (1976). *Human territories. How we behave in space-time.* Englewood Cliffs, NJ: Prentice-Hall Inc.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*, 211-232.

Stevens, J., & Zhang, Y. (2013). Relative distance and gaze in the use of entity-referring spatial demonstratives: An event-related potential study. *Journal of Neurolinguistics*, *26*, 31-45.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*, 705-722.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological review*, *117*, 440-463.

Weinrich, H. (1988:1995). Über Sprache, Leib, und Gedächtnis. In H.-U. Gumbrecht (Ed.), *Materialität der Kommunikation* (pp. 80-93). Frankfurt am Main: Suhrkamp.

Weissenborn, J., & Klein, W. (Eds.). (1982). *Here and there: Cross-linguistic studies on deixis and demonstration*. Amsterdam: John Benjamins.

Chapter 4

Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech

Based on: Peeters, D., Chu, M., Holler, J., Özyürek, A., & Hagoort, P. (2013). Getting to the point: The influence of communicative intent on the kinematics of pointing gestures. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 1127-1132). Austin, TX : Cognitive Science Society.

and

Peeters, D., Chu, M., Holler, J., Hagoort, P. & Özyürek, A. (under review). Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech.

Abstract

In everyday human communication, we often express our communicative intentions by manually pointing out referents in the material world around us to an addressee, often in tight synchronization with referential speech. The present study investigated whether and how the kinematic form of index-finger pointing gestures is shaped by the gesturer's communicative intentions, and how this is modulated by the presence of concurrently produced speech. Furthermore we explored the neural mechanisms underpinning the planning of communicative pointing gestures and speech. Two experiments were carried out in which participants pointed at referents for an addressee while the informativeness of their gestures and speech was varied. Kinematic and electrophysiological data were recorded online. It was found that participants prolonged the duration of the stroke and post-stroke hold phase of their gesture in order to be more communicative, in particular when the gesture was carrying the main informational burden in their multimodal utterance. Frontal and P300 effects in the event-related potentials suggested the importance of intentional and modality-independent attentional mechanisms during the planning phase of informative pointing gestures. These findings contribute to a better understanding of the complex interplay between action, attention, intention, and language in the production of pointing gestures, a communicative act core to human interaction.

Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech

> "... since the motions of the body are obvious and external while those of the soul are invisible and hidden"
> (Petrarch 1336/1898, p.312)

Human communication in everyday life is canonically driven by a speaker's communicative intentions to convey meaning to an addressee, and dependent on the successful recognition of such intentions by that addressee (Bara, 2010; Grice, 1975; Sperber & Wilson, 1995). Ontogenetically, one of the first ways in which we express our communicative intent is by producing pointing gestures to the things around us (e.g., Carpenter, Nagell, & Tomasello, 1998). Such pointing gestures are a foundational building block of human communication (Kita, 2003) and pave the way for the acquisition of language (Bates, Camaioni, & Volterra, 1975; Butterworth, 2003; Csibra, 2010; Moore & D'Entremont, 2001; Iverson & Goldin-Meadow, 2005). Throughout life, in concert with speech, they allow us to directly connect our communication to the material world around us (Bangerter, 2004; Clark, 2003; Enfield, Kita, & De Ruiter, 2007; Kita, 2003). Cognitive models of speech and gesture production generally acknowledge the role of one's communicative intentions in driving the production of co-speech gesture (De Ruiter, 2000; Kita & Özyürek, 2003; Melinger & Levelt, 2004; but see Krauss, Chen, & Gottesman, 2000). Here we investigate whether and how the kinematics of pointing gestures are indeed shaped by one's communicative intentions, and whether this is modulated by the presence of concurrently produced speech. In addition, we explore the neural and cognitive mechanisms involved in the planning of intentionally communicative pointing gestures and speech. To set the stage for a description of two experiments, we will first discuss previous research on communicative actions and intentions in general, and then pointing gestures and speech more specifically.

In everyday life, our hands and arms rarely rest. As humans, we interact with the world around us through manipulating and acting upon objects, and on many occasions, we do not do so just by ourselves but in the context of joint activities involving the presence of others (e.g., Vesper & Richardson, 2014). Crucially, those others have been shown to influence the way we perform instrumental actions (see Becchio, Manera, Sartori, Cavallo, & Castiello, 2012). For example, the movement kinematics of actions such as reaching for and grasping an object have been found to be shaped by the actor's communicative intentions (Sartori, Becchio, Bara, & Castiello, 2009). In turn, observers may derive and anticipate the actor's intentions by attuning to such subtle kinematic parameters in the actor's movement (Sartori, Becchio, & Castiello, 2011). Research in the domain of (representational) co-speech hand gestures (e.g., Özyürek, 2002; see also Gerwing & Bavelas, 2004; Holler & Stevens, 2007) and interpersonal signaling in communicative actions more broadly (Vesper & Richardson, 2014) suggest that the close link between action and intention may not be restricted to instrumental or representational actions (see Pierno, Tubaldi, Turella, Grossi, Barachino, et al., 2009). Preliminary indications indeed suggest a relation between the kinematics of pointing gestures and the speaker-gesturer's communicative intent (Cleret de Langavant, Remy, Trinkler, McIntyre, Dupoux et al., 2011; Enfield et al., 2007). The fieldwork by Enfield and colleagues (2007), for instance, suggests that the size of a pointing gesture may depend on whether it is intended to carry informationally foregrounded or backgrounded information in the speaker's utterance. However, whether, and if so how, the kinematics of pointing gestures, like instrumental actions, are shaped by context-specific communicative intentions remains largely unclear.

Pointing gestures often come with concurrent deictic speech such as spatial demonstratives (e.g., "this" and "that" in English). Speech and gesture are temporally tightly interconnected in the production of referring expressions (e.g., Chu & Hagoort, 2014; Kendon, 2004; Levelt, Richardson, & La Heij, 1985; McNeill, 1992) and can be used independently or simultaneously to single out a referent

(Bangerter, 2004), i.e. an object, person, or event on which one wishes to focus the attention of one's addressee by referring to it. Previous work has investigated whether the presence of speech as a second modality changes the kinematics of a corresponding gesture. Chieffi, Secchi, and Gentilucci (2009) found no kinematic difference between a condition in which participants manually pointed to a remote referent and a condition in which they did the same but also concomitantly produced congruent deictic speech ('there'). In contrast, Gonseth, Vilain, and Vilain (2013) found that pointing gestures produced without corresponding speech had a lower velocity and a longer post-stroke hold phase compared to when deictic speech was concomitantly produced. This discrepancy in findings asks for further investigation.

The current study also aims to advance our understanding of the neural mechanisms involved in the planning and production of pointing gestures. Both in infants and adults, frontal markers of neuronal activity have been identified as being involved in the production of pointing gestures establishing a joint, interpersonal focus of attention on a referent (Cleret de Langavant et al., 2011; Henderson, Yoder, Yale, & McDuffie, 2002; Mundy, Card, & Fox, 2000). This frontal activation has been interpreted as reflecting the involvement of intention-related 'mentalizing' networks (e.g., Brunetti, Zappasodi, Marzetti, Perrucci, Cirillo, et al., 2014). Using magnetoencephalography (MEG), Brunetti et al. (2014) found enhanced activity in dorsal regions of the anterior cingulate cortex (ACC – in medial frontal cortex) for declarative pointing ("pointing to share attention to an object - interpersonal", in their manipulation) compared to imperative pointing ("pointing to request an object - instrumental" in their manipulation), and argue that this difference reflects enhanced mentalizing activity. Central to the difference between the two conditions is the explicit assumption that imperative pointing has only an instrumental purpose. This is problematic though, because arguably also in imperative pointing the person gesturing considers her addressee as a mental, intentional agent when requesting an object by pointing (see Southgate, Van Maanen, & Csibra, 2007). Therefore in the current study we compare two

situations that are both communicative and differ only in the communicative intent of the speaker-gesturer. Furthermore, it is an open question whether also other (e.g., attentional) neuronal mechanisms are involved in the planning and production of communicative pointing, and whether (and if so, how) the presence of concomitantly produced speech interacts with possible intentional and attentional mechanisms involved.

In the current study, we adapted a paradigm introduced by Levelt et al. (1985) in which participants produce pointing gestures in an experimental setting (see also Chu & Hagoort, 2014; De Ruiter, 1998). In our manipulation, participants were asked to point with their index-finger at one of four circles that lit up on a screen with or without producing concurrent speech. Index-finger kinematics, speech, and electroencephalogram (EEG) were continuously recorded. Crucially, as a proxy of the participants' communicative intent in the current study, we manipulated the *informativeness* of the pointing gestures. The notion of informativeness has been used successfully in previous studies to tap into communicative intentions involved in speech production (e.g., Willems, De Boer, De Ruiter, Noordzij, Hagoort, & Toni, 2010). Everyday pointing gestures canonically occur in a context in which interlocutors share a joint attentional frame in which one person directs the attention of another person towards a location, event, or other entity in the perceptual environment, usually precisely to be informative about these referents (Tomasello et al., 2007). In the current study, in line with findings on communicative actions more broadly (Vesper & Richardson, 2014), participants may alter the kinematic properties of their movements in order to make them more informative, for instance by slowing down the movements. Alternatively, different intentions may lead to different patterns of neural activity (see below), but lack behavioral consequences as reflected in the kinematic properties of the pointing movements (cf. Brunetti et al., 2014).

The current approach allows for time-locking event-related potentials (ERPs) not only to the onset of the gesture, but also to the presentation of the stimulus/referent. Several effects can be predicted

on the basis of previous work. Potential frontal effects in the current study may reflect participants' communicative intentions in planning their pointing gestures (cf. Brunetti et al., 2014; Cleret de Langavant et al., 2011; Henderson et al., 2002). Furthermore, upon the intention to produce a more informative gesture, participants may allocate more attentional resources to the task. P3b amplitude may be modulated by task-related cognitive demands that drive attentional resource allocation, such that its amplitude is smaller when a task requires greater amounts of attentional resources (Polich, 2007), in particular when attentional resource allocation is under voluntary control and perceptual quality of the stimuli does not differ across conditions (Kok, 2001), as in our set-up. Smaller amplitude of the stimulus-locked P3b in our study may therefore index that participants voluntarily allocate more attentional resources when planning a more informative gesture for their addressee. A final possibility is that the readiness potential (or 'Bereitschaftspotential'; Kornhuber & Deecke, 1965) is sensitive to our manipulation of communicative intent, which would be marked directly preceding the onset of the pointing gesture's execution as measured over contra-lateral, central electrode sites. Hence, in addition to investigating the effects of communicative intent on pointing gesture production, we also consider specific ERP components during the course of planning informative pointing gestures, including the P3b.

We present two experiments that aim to further our understanding of the basic human communicative act of producing pointing gestures to a visible referent. On the basis of the theoretical considerations outlined above, Experiment 1 investigates i) whether and how communicative intentions shape the kinematic properties of manual pointing gestures, ii) whether and how this is modulated by the presence of speech as a second modality, and iii) the neural mechanisms underlying the communicative intent involved in planning pointing gestures and speech. In everyday multi-modal referential communicative acts, the informational burden can be distributed differentially over the spoken and gestural modalities (e.g., Enfield et al., 2007). Therefore, Experiment 2 tests to what extent the

kinematic and electrophysiological findings obtained in Experiment 1 are modality-independent, i.e. whether they generalize to situations in which speech, rather than gesture, carries the informational burden in identifying a referent for an addressee in a multimodal utterance.

Experiment 1

Method

Participants

Twenty-four native speakers of Dutch (12 female; mean age 20.6), studying at Radboud University Nijmegen, participated in the experiment. They were all right-handed as assessed by a Dutch translation of the Edinburgh Inventory for hand dominance (Oldfield, 1971). Data from two additional participants were discarded due to a large number of trials that contained movement artifacts. Participants had normal or corrected-to-normal vision, no language or hearing impairments or history of neurological disease. They provided written informed consent and were paid 20€ for participation.

Experimental Design and Set-up

Participants were seated at a distance of 100 cm from a computer screen that was placed back-to-back with another computer screen (henceforth: the back screen). Stimuli were four white circles in a horizontal line on the top of the screen the participant viewed, mirroring four circles on the back screen. The circles could light up in either blue or yellow. A second participant (a confederate; henceforth: the addressee) looked at the back screen and the participant's pointing gesture via a camera. Figure 1 shows the addressee's view via the camera (converted to a grayscale image). On all trials, participants referred to the circle that lit up. The addressee noted on a paper form which of the four circles the participant referred to on each trial (in speech and/or gesture). In order to make the deictic act more informative in one case but less informative in the other, the following set-up was used. In both conditions, via a

camera, the addressee observed the pointing gesture of the participant, as well as the circles on the back screen providing the corresponding view of the four circles the participant was seeing. This way, the addressee saw which of the four circles the participant pointed at. Before the arrival of the addressee, the experimenter showed the participant the computer to be used by the addressee and demonstrated that the addressee could see the participant's pointing gestures referring to circles on the computer screen. In this way, the participant knew that the addressee would look at the participant's gestures and to the circles presented on the back screen.

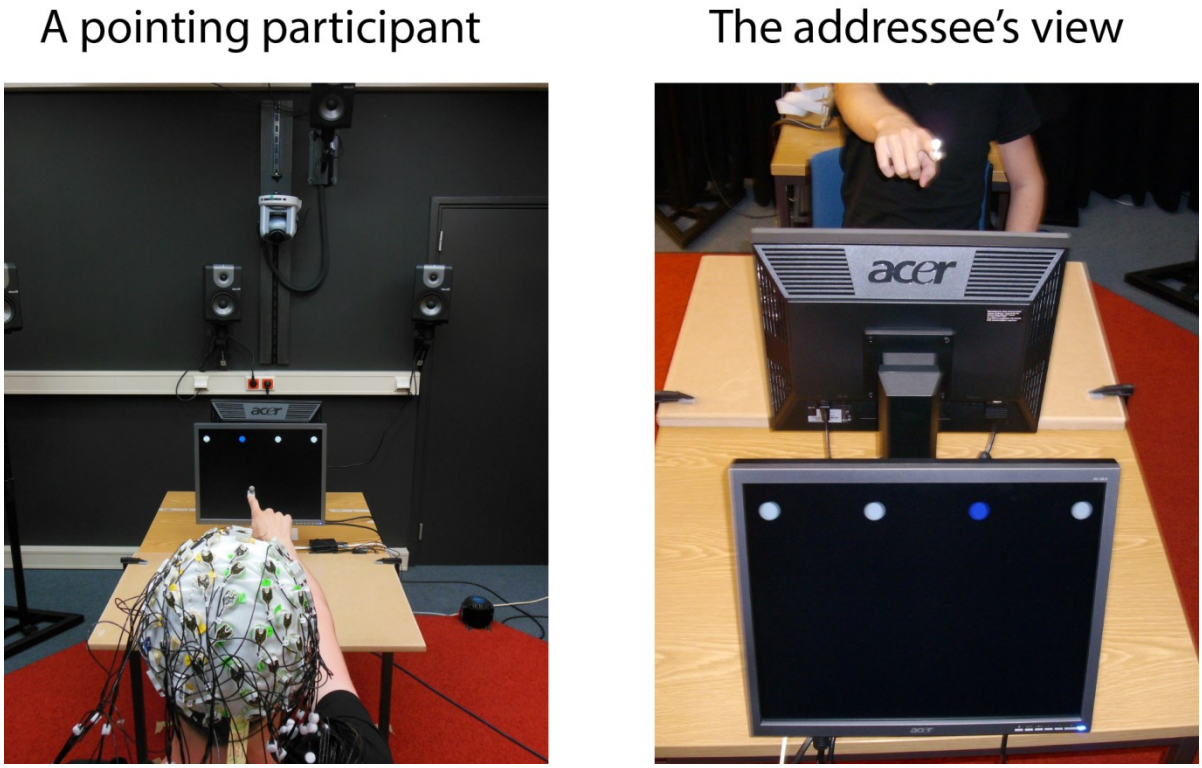

Figure 1. *Left panel: A participant pointing at a circle while EEG, motion tracking kinematics, and speech were continuously recorded. Right panel: The addressee's view of the back screen and the pointing participant during a less informative trial.*

We manipulated the informativeness of the gesture (More Informative versus Less Informative) as well as the modality of the deictic act (Gesture-only versus Gesture + Speech) in a 2x2 within

participants design. In the *more informative condition*, a circle turned blue or yellow only on the participant's screen but not on the back screen. Therefore the participant's pointing gesture was the only source of information on which the addressee could base his/her decision in selecting the circle referred to by the participant. In the *less informative condition*, the respective circle would light up on both the participant's and the addressee's screen. Thus, the participant's pointing gesture was less informative, because the addressee saw the respective circle light up on the back screen at the same moment as the participant saw the corresponding circle light up (i.e., even before the onset of the participant's pointing gesture and/or speech). The participant received written instructions on the screen before each block, specifying whether during that block the addressee would or would not also see circles light up during that block. We decided to not have the addressee give feedback to the participant during the experiment and keep the gesturer's head out of the camera's shot to avoid differences in feedback across conditions and participants (cf. Campisi & Özyürek, 2014; Holler & Wilkin, 2011) and control for the deictic function of eye gaze.

The modality factor was manipulated by having participants use either one or two modalities in referring to the circles. In gesture only blocks (G-only), participants pointed to a circle when it turned blue or yellow without producing speech. In gesture + speech blocks (G+S) participants pointed to the circle and said either *die blauwe cirkel* ("that blue circle") or *die gele cirkel* ("that yellow circle"), depending on the color of the circle. Note that, because any of the four circles could turn blue or yellow on any trial, the speech, which only ever referred to color but never to location, was never informative (neither in the more informative nor the less informative blocks) in this Experiment. The rationale for this was that we were interested in the possible effect of the mere presence of speech as a second modality, in addition to the informativeness of the deictic act that was manipulated separately in the gesture. Figure 2 gives an overview of the manipulation.

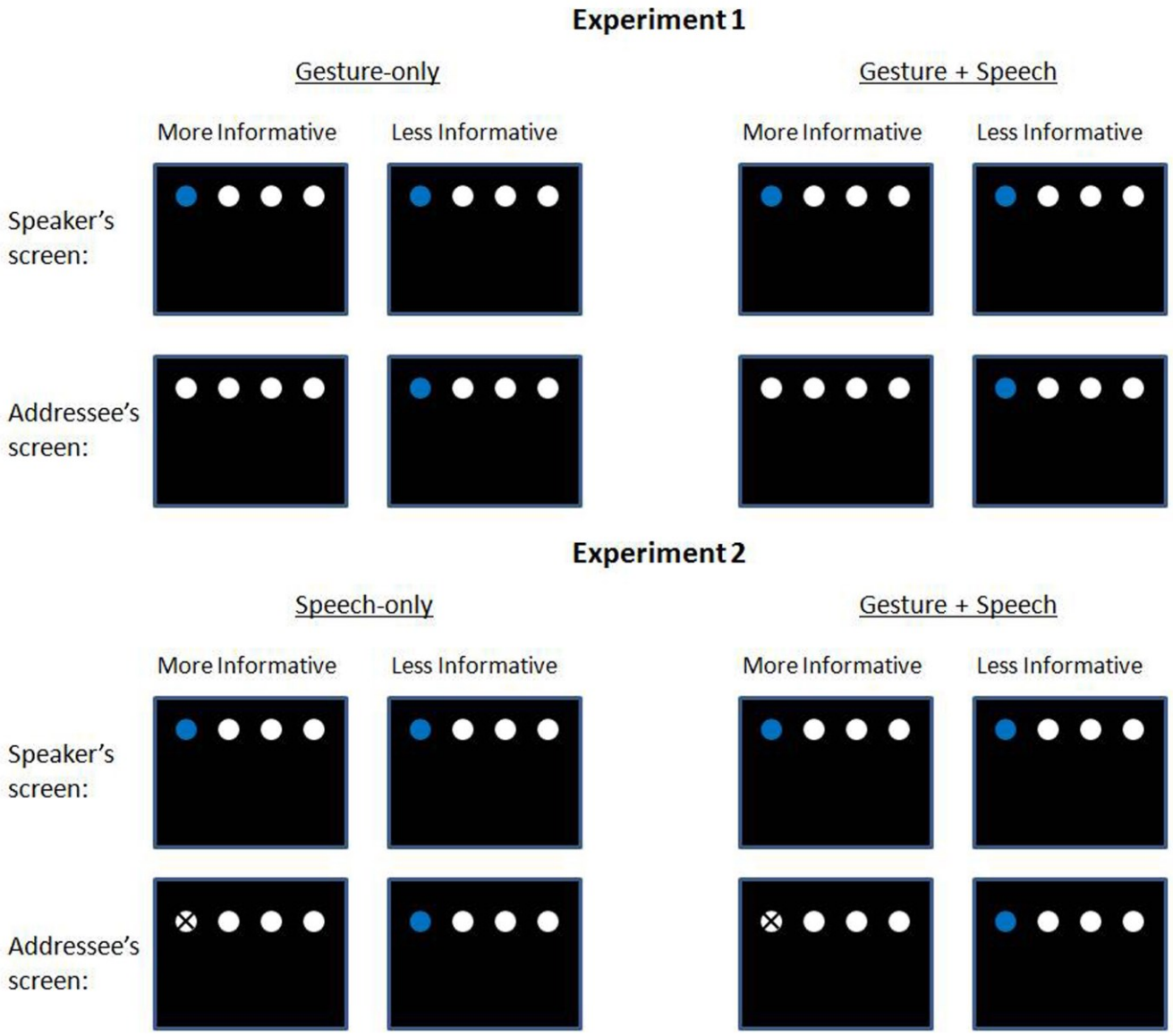

Figure 2. *Overview of the design of the Experiments 1 and 2.*

Each trial started with a fixation cross, displayed for 500 ms, followed by the presentation of four white circles. After a jittered period of 500-1000 ms, one of the circles turned yellow or blue. At this point, the participant was allowed to release her finger from a button, pointed to the blue or yellow circle, and (in the G+S blocks), in speech, referred to the color of the circle. The experiment consisted of

16 blocks of 20 trials each. Every condition in the experiment was represented by four blocks. The order of presentation of blocks was counterbalanced across participants. In half of the trials a circle lit up yellow, in the other half it lit up blue. Each block of 20 trials consisted of ten circles lighting up yellow and ten lighting up blue, equally distributed over the four circles and the four conditions throughout the experiment, in a randomized way.

Procedure

On arrival of the participant, the experimenter explained that a second participant (i.e., the confederate addressee) would perform a behavioral task on the basis of the participant's gestures. The experimenter showed the participant the computer and form to be used by the addressee and demonstrated that the addressee could view the participant's pointing gestures referring to circles on the computer screen.

In order to keep participants motivated, it was emphasized that they were in a joint activity with the addressee and that the success of this joint activity depended on how well they worked together. The participant was then seated in a comfortable chair in the experiment room. The height of the screen was adjusted to the height of the eyes of the participant. The button used by the participant was placed at the height of the participant's elbow, 23 cm in front of the participant calculated from the vertical axis corresponding to the position of the participant's eyes. Participants were instructed to always rest their finger on this button, except when making the pointing gesture, which allowed calculating the duration and onset of the pointing gesture. An active, wireless sensor was placed on the participant's right index finger nail to allow for motion tracking of the pointing movements. Participants' electroencephalogram (EEG) was recorded continuously throughout the experiment (see below).

After montage of the motion tracking sensor the experimenter picked up the confederate addressee. The addressee was shown the room in which the participant performed the task, greeted the

participant, and was seated in a chair in front of a computer in a room adjacent to the participant's room. Thirty-two test-items (eight per condition) preceded the main experiment as a practice set. Participants received specific instructions to point with or without speech before each block. In addition, before each block, the participant was instructed whether the addressee could also see the same circles light up at the back screen or not during that block. Participants were asked to only move their hand and arm when pointing. During the experiment, participants were allowed to have a short break after every fourth block. Before and during the experiment, the communication between experimenter and addressee was minimal and fully scripted, in order to be consistent across participants. The addressee provided no feedback to the participant during the experiment. After the experiment, the addressee was thanked for participation and left the room. After filling out a post-test questionnaire, participants were debriefed, financially compensated, and thanked for participation. The results of the post-test questionnaire revealed that all participants thought the confederate addressee was another (naive) participant who performed well on his task.

Kinematic and speech recording and analysis

Behavioral and kinematic data were acquired throughout the experiment using experimental software (Presentation, Neurobehavioral Systems, Inc) and a 60 Hz motion tracking system and DTrack2 tracking software (both Advanced Realtime Tracking, Weilheim, Germany). In line with previous work (Chu & Hagoort, 2014; Levelt et al., 1985) we focused on different kinematic aspects of the pointing movements, including the gesture initiation time, the stroke duration, the apex time, the hold duration, the incremental distance traveled by the pointing finger, and the velocity of the movement. *Praat* software (version 5.2.46; Boersma & Weenink, 2009) was used to calculate offline the speech duration, and the maximum loudness and mean loudness of speech. Table 1 gives an overview of

how the kinematic and speech-related dependent variables were defined and calculated (cf. Chu & Hagoort, 2014; Levelt et al., 1985).

Electrophysiological recording and analysis

Throughout the experiment, the participant's electroencephalogram (EEG) was recorded continuously from 59 active electrodes (Brain Products, Munich, Germany) held in place on the scalp by an elastic cap (Neuroscan, Singen, Germany). In addition to the 59 scalp sites, three external electrodes were attached to record participants electrooculogram (EOG), one below the left eye (to monitor for vertical eye movement/blinks), and two on the lateral canthi next to the left and right eye (to monitor for horizontal eye movements). Finally, one electrode was placed over the left mastoid bone and one over the right mastoid bone. All electrode impedances were kept below 20 KΩ. The continuous EEG was recorded with a sampling rate of 500 Hz, a low cut-off filter of 0.01 Hz and a high cut-off filter of 200 Hz. EEG was filtered offline (high-pass at 0.01 Hz and low-pass at 40 Hz). All electrode sites were referenced online to the electrode placed over the left mastoid and re-referenced offline to the average of the right and left mastoids.

Markers were sent from the computer presenting the stimuli to the computer recording the EEG, at light onset and at gesture initiation. Using Brain Vision Analyzer software (Brain Products, Munich, Germany), event-related potentials (ERPs) were time-locked to light onset (i.e., stimulus-locked) and to gesture initiation (i.e., the onset of the pointing gesture; henceforth called "gesture-locked"). In the stimulus-locked ERPs, the 100 ms pre-stimulus period was used as a baseline. In the gesture-locked ERPs, the period 700 to 600 ms before gesture initiation was used as a baseline, because this time-window reliably preceded stimulus onset (see Gesture Initiation Time in Table 2), such that the gesture-locked ERP would globally reflect the time between stimulus onset and gesture initiation. Note that in both the stimulus-locked and the gesture-locked ERPs we thus look at the activity preceding the onset of

Table 1. *Definition of the (behavioral) kinematic and speech-related dependent variables in Experiments 1 and 2, as calculated for each experimental trial.*

Variable	Definition
Kinematic Dependent Variables	
Gesture Initiation Time (ms)	Gesture Onset – Light Onset
Stroke Duration (ms)	Gesture Apex – Gesture Onset
Apex Time (ms)	Gesture Apex – Light Onset
Hold Duration (ms)	Retraction Time – Gesture Apex
Incremental Distance (cm)	The amount of distance travelled by the pointing index-finger between Gesture Onset and Gesture Apex
Velocity (cm/s)	Apex Time / Incremental Distance
Speech-related Dependent Variables	
Speech Duration (ms)	Speech Offset – Speech Onset
Speech Onset Time (ms)	Speech Onset – Light Onset
Synchronization Time (ms)	Speech Onset Time – Apex Time
Maximum Loudness (db)	The maximum loudness of speech during an utterance
Mean Loudness (db)	The average loudness of speech across an utterance
Other variables used in calculations above	
Light Onset	The moment in time a circle lit up
Gesture Onset	The moment in time the participant's finger left the button in order to point
Gesture Apex	The moment in time where the pointing index-finger was at least 7 cm from the button and moved forward only less than 2 mm for two consecutive samples
Speech Onset	The moment in time the participant started speaking
Speech OffSet	The moment in time the participant stopped speaking
Retraction Time	The moment in time where the pointing index-finger moved back in the direction of the button for at least 2 mm in two consecutive samples

the gesture. Trials containing muscular artifacts were removed from further analysis (5.5% of the total stimulus-locked dataset; 13.7% of the total gesture-locked dataset). The amount of removed trials was similar across the different levels of the Informativeness and Modality factors. Subsequently, ICA was used to correct for ocular artifacts (extended infomax procedure, cf. Lee, Girolami, & Sejnowski, 1999). The mean amplitudes of the ERP waveforms for each condition per subject were entered into repeated measures ANOVAs in a time-window analysis of 100-ms time windows after stimulus onset (0-400 ms) or before gesture initiation (-600 ms until gesture onset) respectively. A subset of five regions of interest was selected for the analyses (see Figure 3) based on previous, related work outlined in the Introduction. An anterior ROI was selected on the basis of the findings in Henderson et al. (2002). A potential modulation of the readiness potential as a function of our Informativeness manipulation would be reflected in an effect over left central but not right central electrode sites because all participants were right-handed and pointing with their right index-finger. Therefore a left middle and a right middle ROI were selected. Finally, a possible P300 (P3b) effect would occur in posterior electrode sites, possibly right-lateralized (Polich, 2007), which led to the selection of a left posterior and a right posterior ROI. In sum, the ERP analyses thus contained the independent variables Informativeness (More Informative versus Less Informative), Modality (Gesture-only versus Gesture+Speech), and Region of Interest (ROI: Anterior, Left Middle, Right Middle, Left Posterior, Right Posterior). The Greenhouse and Geisser (1959) correction was applied when appropriate. Corrected degrees of freedom are reported.

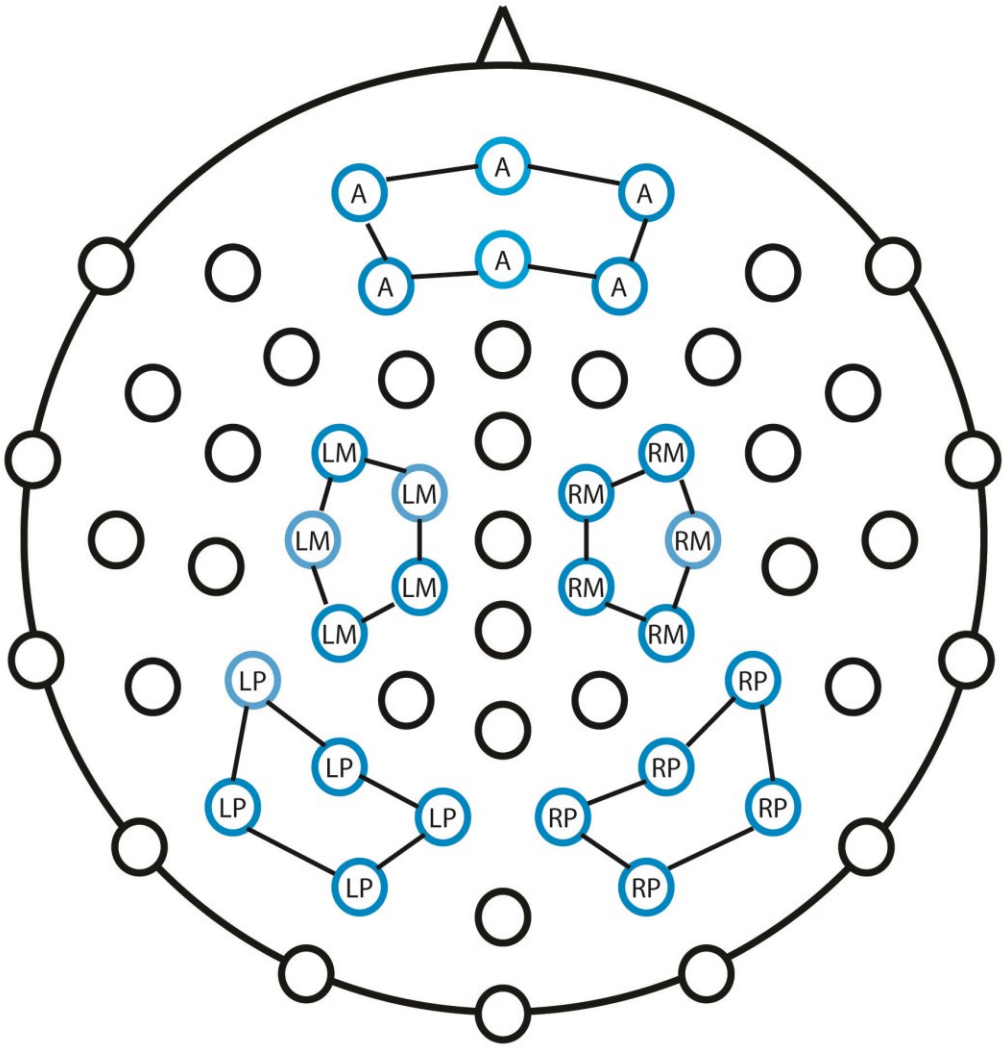

Figure 3. *Electrode montage. Five regions of interest were used in the analysis of the electrophysiological data: anterior (A), left middle (LM), right middle (RM), left posterior (LP), and right posterior (RP).*

Results

Behavioral Results

Trials on which the Gesture Initiation Time was below 100 ms or above 2000 ms were considered errors and excluded from all analyses (0.7% of total dataset). In addition, trials containing

hesitations or errors in the participant's speech were removed from further analyses (0.2% of all data). Separate analyses of variance were performed for each dependent variable with Informativeness (More Informative versus Less Informative) and Modality (Gesture-only versus Gesture + Speech) as within-subject factors. The analyses performed on the Gesture Initiation Time and the Incremental Distance did not yield any significant main or interaction effects (all p's >.05).

The analysis of the Stroke Duration yielded a significant main effect of Informativeness, F (1,23) = 10.97, p = .003, η_p^2 = .32. This effect reflected that the duration of the stroke was significantly longer in the More Informative condition (M = 837 ms) than in the Less Informative condition (M = 823 ms). No significant main effect of Modality was found. There was no significant interaction between the two factors.

The analysis of the Apex Time showed a significant main effect of Informativeness, F (1,23) = 8.15, p = .009, η_p^2 = .26. This effect denoted that the apex was reached significantly later in the More Informative condition (M = 1379 ms) than in the Less Informative condition (M = 1359 ms). No significant main effect of Modality was found. There was no significant interaction between the two factors.

The analysis on the mean Velocity yielded a significant main effect of Informativeness, F (1,23) = 5.75, p = .025, η_p^2 = .20. The velocity of the pointing gesture was significantly lower in the More Informative condition (M = 38.2 cm/s) than in the Less Informative condition (M = 38.7 cm/s). Again, no significant main effect of Modality or interaction between the two factors was found.

The analysis performed on the Hold Duration yielded a significant main effect of Informativeness, F (1,23) = 10.17, p = .004, η_p^2 = .31. The Hold Duration was significantly longer in the More Informative condition (M = 1235 ms) compared to the Less Informative condition (M = 1143 ms). No significant main effect of Modality was found and there was no significant interaction between the two factors.

An analysis on the Speech Onset Time (G+S conditions only) revealed a significant main effect of Informativeness, F (1,23) = 6.79, p = .016, η_p^2 = .23. This effect reflected that the speech onset on average took place significantly later in the More Informative condition (M = 1385 ms) than in the Less Informative condition (M = 1351 ms).

Finally, an analysis on the Synchronization Time in the G+S conditions did not show a significant main effect of Informativeness (p = .16), indicating that the onset of the speech and the apex of the gesture were aligned similarly across conditions and independently from the informativeness of the gesture. The Maximum Loudness and Mean Loudness of the speech did not differ significantly across Informativeness, nor did the Speech Duration (all p's >.05).

In sum, participants prolonged the duration of the stroke and the hold-phase of their pointing gesture in order to be more informative, which led to a gesture with a lower velocity and delayed the moment at which the apex was reached. Table 2 summarizes all behavioral results from Experiment 1.

Electrophysiological results

Trials defined as errors or outliers in the behavioral analyses were also excluded from the ERP analyses. Separate ERPs were computed for each condition in the experiment. By-participant analyses (both stimulus-locked and gesture-locked) were performed with Informativeness, Modality, and ROI as independent variables. Only significant effects at the 5% level are reported, unless explicitly stated otherwise.

Table 2. *Overview of the behavioral results per condition in Experiment 1. Duration in ms is displayed for Gesture Initiation Time (GIT), Stroke Duration (Stroke), Apex Time (Apex), Hold Duration (Hold), Speech Duration (SpeechDur), Speech Onset Time (SOT), and Synchronization Time (Sync). Further, the Incremental Distance in cm (Dist), Velocity in cm/s (Velocity), and the maximum and mean loudness of the speech (Max_Loudness and Mean_Loudness) in db are provided. The standard error of the mean is indicated between parentheses. An asterisk next to a variable's name indicates a significant main effect of Informativeness in the analysis.*

Condition	GIT	Stroke*	Apex*	Dist	Velocity*	Hold*
More Informative						
Gesture-only	534 (21)	834 (30)	1368 (42)	51 (1)	38.5 (1)	1252 (135)
Gesture + Speech	550 (22)	840 (27)	1389 (39)	51 (1)	37.8 (1)	1219 (121)
Less Informative						
Gesture-only	532 (22)	819 (29)	1351 (41)	51 (1)	39.0 (1)	1138 (116)
Gesture + Speech	541 (24)	826 (27)	1367 (40)	51 (1)	38.5 (1)	1149 (106)

Condition	SpeechDur	SOT*	Sync	Max_Loudness	Mean_Loudness
More Informative					
Gesture-only					
Gesture + Speech	1167 (35)	1385 (65)	4 (54)	82.0 (1)	70.8 (1)
Less Informative					
Gesture-only					
Gesture + Speech	1155 (36)	1351 (66)	16 (54)	82.2 (1)	70.8 (1)

Stimulus-locked analysis. The omnibus stimulus-locked analysis firstly revealed a significant Informativeness x ROI interaction effect in time-windows 200-300 ms, F (2,52) = 4.58, p = .012, η_p^2 = .17, and 300-400 ms, F (2,45) = 3.39, p = .044, η_p^2 = .13 after stimulus onset. Follow-up analyses showed a significant main effect of Informativeness in the 300-400 ms time-window in the right posterior ROI only, F (1,23) = 5.53, p = .028, η_p^2 = .19. This effect reflected a significantly more positive ERP wave for the More Informative condition compared to the Less Informative condition. We will refer to this effect as a P300 or P3b effect (cf. Polich, 2007). There was a trend towards a similar effect of Informativeness in the 200-300 ms time-window in the right posterior ROI, F (1,23) = 3.14, p = .090, η_p^2 = .12.

Second, the omnibus analysis revealed a significant main effect of Modality in the 100-200 ms time-window, F (1,23) = 6.27, p = .020, η_p^2 = .21, the 200-300 ms time-window, F (1,23) = 4.77, p = .039, η_p^2 = .17, and the 300-400 ms time-window, F (1,23) = 11.17, p = .003, η_p^2 = .33. These main effects of Modality reflected a significantly more positive ERP wave for the Gesture+Speech condition compared to the Gesture-only condition. No Informativeness x Modality interaction effect was found in any time-window (all F's $\leq$ 1). Figure 4 graphically presents the stimulus-locked ERP results.

Gesture-locked analysis. The omnibus analysis locked to the onset of the gesture revealed a significant Informativeness x ROI interaction effect in the -100-0 ms time-window, i.e. directly preceding gesture initiation, F (3,66) = 6.09, p = .001, η_p^2 = .21. Follow-up analyses yielded a significant main effect of Informativeness in this time-window in the anterior ROI only, F (1,23) = 5.03, p = .035, η_p^2 = .18. This effect reflected a significantly more negative ERP-wave for the Less Informative condition compared to the More Informative condition (see Figure 5). We will refer to this effect as a frontal marker of informativeness / communicative intent. No such effect was found in any

other ROI (all $F's \leq 1$). No main effect of Modality or Informativeness x Modality interaction effects were found (all $F's < 1$).

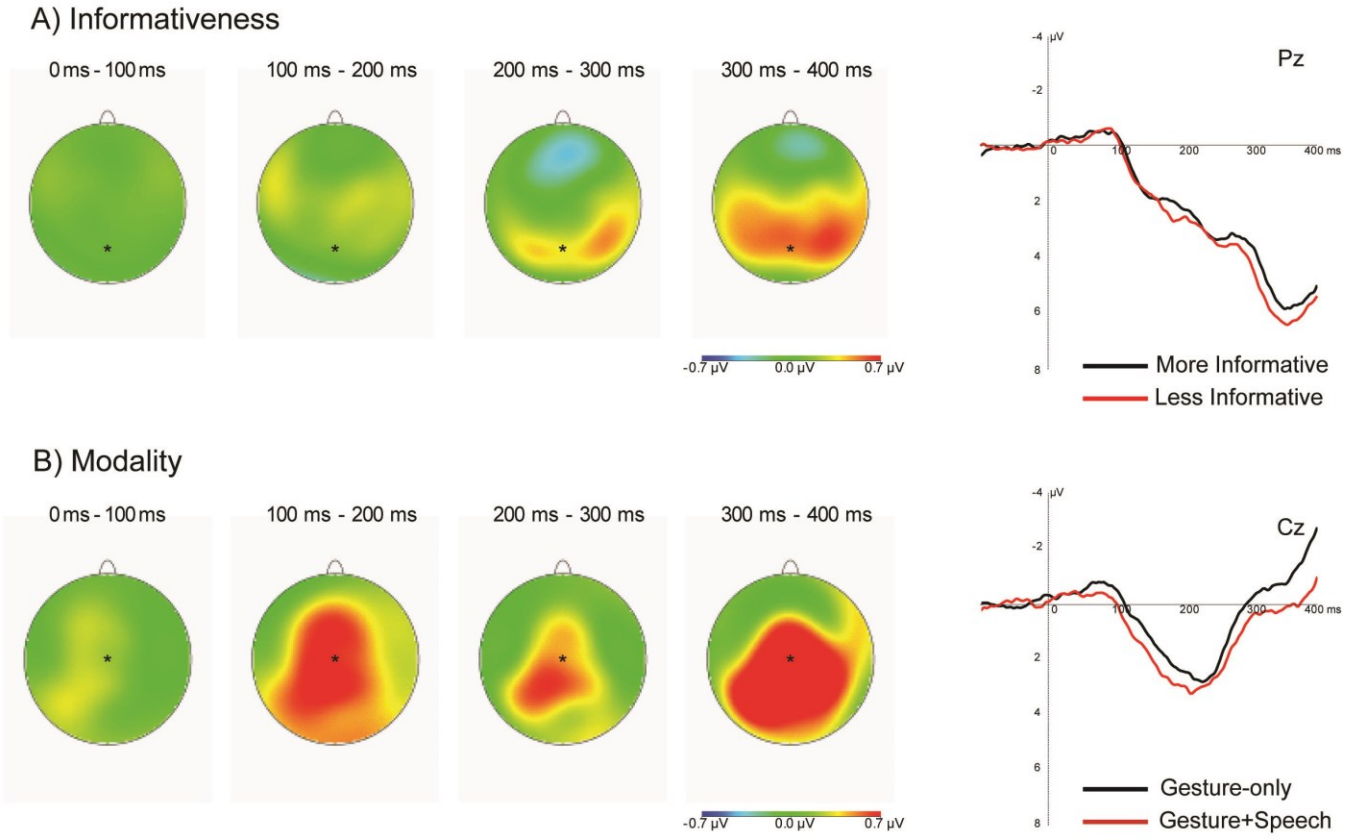

Figure 4. *Grand average waveforms and topographic plots corresponding to the voltage difference between conditions in subsequent time-windows in the stimulus-locked ERP analysis in Experiment 1 for A) The main effect of Informativeness (collapsed over Modality) and B) the main effect of Modality (collapsed over Informativeness).The electrode site used for the waveforms is indicated in the corresponding topographic plots.*

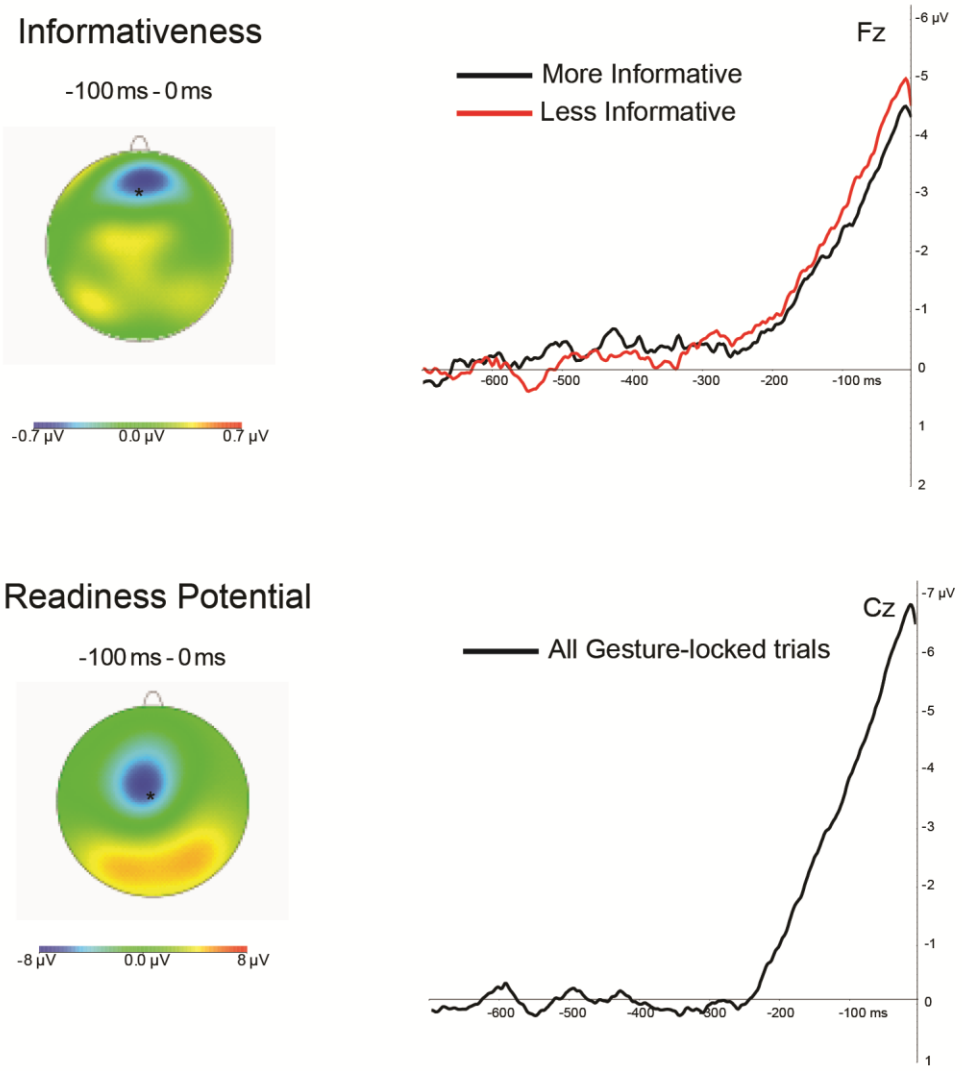

Figure 5. *Top panel: Grand average waveforms and topographic plot corresponding to the voltage difference between conditions for the main effect of Informativeness (collapsed over Modality) in the gesture-locked ERP analysis in Experiment 1 in the time-window directly preceding gesture initiation in the anterior column. Bottom panel: The readiness potential and its locus over the scalp across all gesture-locked trials. The electrode site used for the waveforms is indicated in the topographic plots.*

Discussion

Experiment 1 revealed behavioral and electrophysiological correlates of communicative intent in the planning and production of index-finger pointing gestures.

Behaviorally, participants prolonged the duration of the stroke of their pointing gesture in order to be more informative, which led to a gesture with a lower velocity and delayed the moment at which the apex was reached. In addition, the post-stroke hold-phase of the gesture was maintained for longer. The kinematic properties of participants' pointing gestures were not affected by the concurrent production of speech (in line with Chieffi et al., 2009) and similar kinematic effects of communicative intent were found in situations where people only used gesture to communicate compared to situations where speech and gesture were simultaneously produced. In addition, participants temporally aligned the onset of their deictic linguistic expression with the moment the pointing gesture reached its apex, regardless of whether the gesture was more or less informative. No effect of participants' communicative intentions was found in the quality of the (largely informationally redundant) speech itself. We will discuss the theoretical implications of these findings in the General Discussion.

Neurophysiologically, the stimulus-locked ERPs showed a (parietal) P3b effect with smaller amplitude for the more informative condition, independent of modality. As outlined in the Introduction, P3b amplitude may be modulated by task-related cognitive demands that drive attentional resource allocation, such that its amplitude is smaller when a task requires greater amounts of attentional resources (Polich, 2007). Smaller amplitude of the stimulus-locked P3b in the more informative condition may therefore reflect that participants voluntarily allocated more attentional resources when planning a more informative gesture for their addressee. Furthermore, the gesture-locked waveforms showed a frontal marker of communicative intent directly preceding the onset of the pointing movement. As shown in Figure 5, they resembled the readiness potential (Kornhuber & Deecke, 1965), but had a clearly different distribution over the scalp (i.e., more anterior and less lateralized). The frontal locus of this effect is reminiscent of the locus of electrophysiological findings in infant studies tapping into developing joint attentional mechanisms related to pointing in infancy (e.g., Henderson et al., 2002). More generally, in the planning and production of pointing gestures, frontal effects have been

interpreted as reflecting the involvement of intention-related 'mentalizing' networks (e.g., Brunetti et al., 2014). Our effect modulating the readiness potential (see Figure 5) thus suggests an interaction between planning a motor program and activation of the mentalizing network (Amodio & Frith, 2006).

It is an open question to what extent the kinematic and electrophysiological findings obtained in Experiment 1 are specific to situations in which the gesture is carrying the main informational burden in a multimodal speech act. It is possible that whenever speech itself is informative enough to single out a referent, people no longer design the kinematics of their concomitant gesture to be maximally informative (Bangerter, 2004; Cooperrider, 2011; see also Enfield, Kita, & De Ruiter, 2007), although they may still have a similar communicative intention. We tested this possibility in a second experiment, presented below, in which participants crucially had to refer to the color of the circle that lit up, and the addressee's task was to note down the color of the circle. Because the color and not the location of the circle was now the important aspect of the stimulus, in this case the speech modality and not the gestural modality carried the informational burden. This manipulation thus allowed us to investigate whether the extent to which people modify the kinematic characteristics of their pointing gesture on the basis of their communicative intentions is dependent on how they distribute the informational burden over two modalities (speech and gesture). In Experiment 2, therefore, the informativeness of the deictic act was thus manipulated in the speech modality, which was either paired with a redundant pointing gesture (bimodal condition) or not (unimodal condition).

Furthermore, Experiment 2 will show whether the intentional and attentional neurophysiological markers that we found in Experiment 1 are specific to cases where pointing gestures carry the main informational burden or whether they are modality-independent instead. A frontal speech-locked ERP effect of Informativeness may suggest a common intention-related mechanism in the planning of both referential gesture and speech. Moreover, if the stimulus-locked P3b effect indeed reflects (voluntary) attentional resource allocation, it will be independent of whether participants' task is to refer to the

spatial location (as in Experiment 1) or color (as in Experiment 2) of the entity they point at and attend to.

Experiment 2

Method

Participants

Twenty-four new participants (12 female; mean age 21.7) matching the criteria from Experiment 1 took part in Experiment 2. Data from six additional participants were obtained but had to be discarded due to technical failure during the experiment or due to the presence of a large number of trials that contained movement artifacts. All participants provided written informed consent and were paid 20€ for participation.

Experimental Design and Set-up

Similar to Experiment 1, stimuli were four white circles in a horizontal line on the top of the screen, mirroring four circles on the back screen. Each circle could light up in blue or yellow. Again, the addressee (the same confederate as in Experiment 1) looked at the back screen (providing the corresponding view of the four circles the participant was seeing) and the actual participant via a camera. On all trials, participants referred to the circle that lit up. In contrast with Experiment 1, the addressee noted on a paper form the *color* of the circle that lit up (and not the *location*). In addition, the addressee listened to the participant's speech via speakers in the addressee's room.

The informativeness of the speech (More Informative versus Less Informative) as well as the modality of the deictic act (Speech-only versus Gesture + Speech) were manipulated in a 2x2 within participants design. In the More Informative condition, a circle turned blue or yellow only on the

participants' screen but not on the back screen. In order to make the pointing gesture in the Speech + Gesture condition redundant, the location of the circle that lit up was marked by a cross in the More Informative condition on the back screen only (see Figure 2). The participant's speech was the only source of information on which the addressee had to base his decision in determining the color of the circle referred to by the participant. In the Less Informative condition, the corresponding circles would light up on both the participant's and the addressee's screen. This rendered the participant's speech less informative in this condition, because the addressee saw the respective circle light up in either blue or yellow on the back screen at the same moment as the participant saw the corresponding, same color circle light up.

The modality factor was manipulated by having participants use either one or two modalities in referring to the circles. In speech only blocks (S-only), when a circle lit up participants said *de blauwe cirkel* ("the blue circle") or *de gele cirkel* ("the yellow circle"), depending on the color of the circle, without producing a pointing gesture. In Gesture + Speech blocks (G+S) participants uttered the same phrase but now also produced an index-finger pointing gesture towards the location of the circle that lit up. Note that, because on all trials the location of the circle was known by the addressee and because the location of the circle was irrelevant for the task performed by the addressee in Experiment 2, the gesture was never informative (neither in the More Informative nor in the Less Informative blocks). The rationale for this was that we were interested in the possible effect of the mere presence of gesture as a second modality, independently from the informativeness of the deictic act that was manipulated separately in the speech modality. The trial structure was the same as in Experiment 1.

Procedure

The experimental procedure was the same as in Experiment 1. Again, the results of the post-test questionnaire revealed that all participants thought the confederate addressee was another (naive) participant who performed well on his task.

Kinematic, electrophysiological, and speech recordings

The kinematic, electrophysiological, and speech recordings were done similarly to Experiment 1. EEG was recorded continuously and ERPs were time-locked separately to light onset (i.e., stimulus-locked), gesture initiation (in the G+S blocks; "gesture-locked"), and voice onset (in the S-only blocks; "speech-locked"). The stimulus-locked preprocessing and ERP analyses were the same as in Experiment 1. The gesture-locked analysis was also the same as in Experiment 1 except for the absence of the Modality factor due to gesture being produced only in G+S blocks in this experiment. An additional analysis in 100-ms time-windows preceding the speech was carried out on ERPs time-locked to speech onset during S-only blocks. Separate analyses were carried out for 100-ms time-windows during the 900 ms preceding speech onset. The 1000 ms to 900 ms time-window preceding speech onset was used as a baseline period, because this time-window reliably preceded speech onset time in the S-only blocks (regardless of Informativeness). Trials containing muscular artifacts were removed from further analysis (7.4% of the total stimulus-locked dataset; 17.1% of the total gesture-locked dataset; 8.2% of the total speech-locked dataset). The amount of removed trials was similar across the different levels of Informativeness and Modality. Inspection of the EEG data confirmed that it was not feasible to further look into speech-locked ERPs in the G+S blocks due to the concurrent pointing gesture creating movement artifacts prior to speech onset (gesture onset systematically preceded voice onset).

Results

Behavioral Results

Trials on which the Gesture Initiation Time or the Speech Onset Time was below 100 ms or above 2000 ms were considered errors and excluded from all analyses (0.5% of total dataset). In addition, trials containing hesitations or errors in the participant's speech were removed from further analysis (0.3% of all data).

First, separate analyses of variance were performed on the Speech Duration and the Speech Onset Time with Informativeness (More Informative versus Less Informative) and Modality (Speech-only or Gesture+Speech) as within-subject factors. The analysis of the Speech Onset Time revealed a significant main effect of Modality, F (1,23) = 87.49, p = .001, η_p^2 = .79, with the speech onset being significantly later in the G+S condition (M = 976 ms) compared to the S-only condition (M = 706 ms). The analysis of the Speech Duration yielded a significant main effect of Modality, F (1,23) = 5.74, p = .025, η_p^2 = .20 driven by the speech duration being significantly longer in the G+S condition (M = 1111 ms) compared to the S-only condition (M = 1095 ms).

Both the analysis of the Maximum Loudness of the speech and the analysis of the Mean Loudness of the speech showed a significant main effect of Modality, F (1,23) = 16.55, p = .001, η_p^2 = .42 and F (1,23) = 8.73, p = .007, η_p^2 = .28 respectively. This indicated that participants spoke louder in the bimodal compared to the unimodal conditions. In all these analyses, no significant main effect of Informativeness was found and there was no significant interaction between the two factors.

In the G+S conditions, participants manually pointed at the circle on the screen while linguistically referring to it. Repeated measures analyses of variance with Informativeness as the single within-subject factor were carried out on the same dependent variables as in Experiment 1. The analysis of the Stroke Duration showed a significant main effect of Informativeness, F (1,23) = 5.42, p = .029, η_p^2 = .19. This effect denoted that the duration of the stroke was significantly longer in the More Informative condition (M = 707 ms) than in the Less Informative condition (M = 698 ms). Analyses of

Table 3. *Overview of the behavioral results per condition in Experiment 2. Duration in ms is displayed for Gesture Initiation Time (GIT), Stroke Duration (Stroke), Apex Time (Apex), Hold Duration (Hold), Speech Duration (SpeechDur), Speech Onset Time (SOT), and Synchronization Time (Sync). Further, the Incremental Distance in cm (Dist), Velocity in cm/s (Velocity), and the maximum and mean loudness of the speech (Max_Loudness and Mean_Loudness) in db are provided. The standard error of the mean is indicated between parentheses. An asterisk next to a variable's name indicates a significant main effect of Informativeness in the analysis.*

Condition	GIT	Stroke*	Apex	Dist	Velocity	Hold
More Informative						
Speech-only						
Gesture + Speech	552 (27)	707 (24)	1259 (39)	42 (1)	34.3 (1)	592 (78)
Less Informative						
Speech-only						
Gesture + Speech	548 (26)	698 (25)	1247 (38)	42 (1)	34.5 (1)	576 (76)

Condition	SpeechDur	SOT	Sync	Max_Loudness	Mean_Loudness
More Informative					
Speech-only	1095 (39)	711 (31)		78.6 (1)	66.1 (1)
Gesture + Speech	1114 (43)	977 (46)	28 (3)	79.5 (1)	66.5 (1)
Less Informative					
Speech-only	1095 (42)	702 (32)		78.5 (1)	65.8 (1)
Gesture + Speech	1108 (44)	976 (48)	27 (3)	79.2 (1)	66.5 (1)

Gesture Initiation Time, Apex Time, Incremental Distance, Velocity, Hold Duration, and Synchronization Time did not yield any significant effect (all *p*'s >.05). Table 3 summarizes all behavioral results from Experiment 2.

Electrophysiological Results

Trials defined as errors or outliers in the behavioral analyses were also excluded from the ERP analyses. Separate ERPs were computed for each condition in the experiment. By-participant analyses were performed with Informativeness and ROI as independent variables. In the stimulus-locked analysis Modality was added as a factor. Only significant effects at the 5% level are reported, unless explicitly stated otherwise.

Stimulus-locked analysis. The omnibus stimulus-locked analysis firstly revealed a significant Informativeness x ROI interaction effect in the 200-300 ms time-window, F (2,42) = 3.32, p = .049, η_p^2 = .13, and in the 300-400 ms time-window, F (2,40) = 4.37, p = .024, η_p^2 = .16. Follow-up analyses revealed that these interactions reflected a significant main effect of Informativeness in the right posterior ROI in the 200-300 ms time-window, F (1,23) = 8.87, p = .007, η_p^2 = .28, and in the 300-400 ms time-window, F (1,23) = 7.19, p = .013, η_p^2 = .24. We will again refer to this effect as a P300 or P3b effect (cf. Polich, 2007). A trend towards such a main effect of Informativeness was found in the left posterior ROI in the 300-400 ms time-window, F (1,23) = 3.68, p = .068, η_p^2 = .14. No main effects of Informativeness were found in the other ROIs (all F's < 1).

Secondly, the omnibus analysis revealed a significant Modality x ROI interaction effect in the 100-200 ms time-window, F (3,59) = 3.11, p = .040, η_p^2 = .12, in the 200-300 ms time-window, F (2,56) = 12.88, p = .001, η_p^2 = .36, and in the 300-400 ms time-window, F (3,66) = 24.73, p = .001, η_p^2 = .52. Follow-up analyses revealed a significant main effect of Modality that reached significance in the 200-300 ms time-window in the left middle ROI only (p < .001) before becoming significant in the middle

ROIs in the 300-400 ms time-window as well (all *p*'s <.01), but not in the two posterior ROIs (both *F*'s <1). Figure 6 shows the effects of Informativeness and Modality in the stimulus-locked analysis.

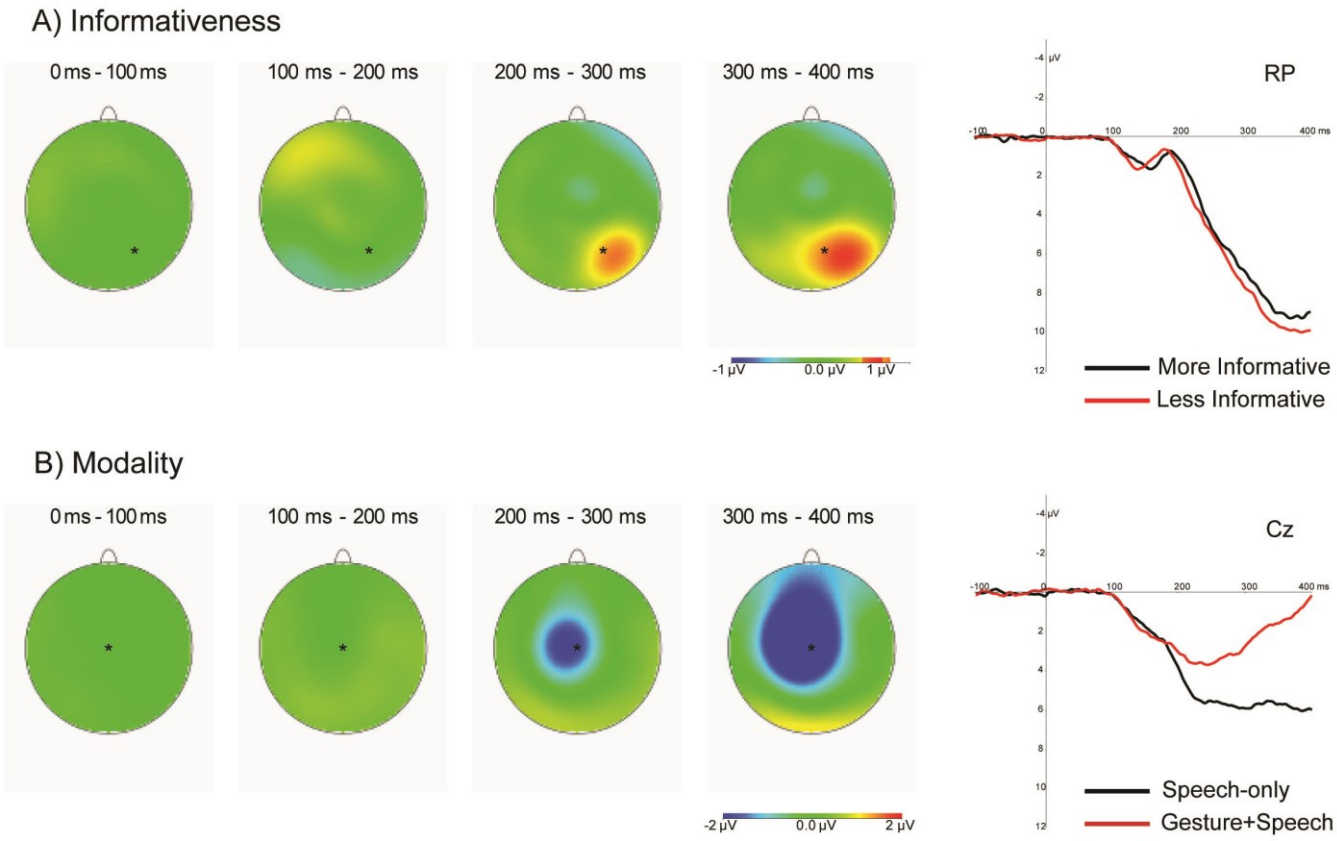

Figure 6. *Grand average waveforms and topographic plots corresponding to the voltage difference between conditions in subsequent time-windows in the stimulus-locked ERP analysis in Experiment 2 for A) The main effect of Informativeness (collapsed over Modality) and B) the main effect of Modality (collapsed over Informativeness). The electrode site used for the waveforms is indicated in the corresponding topographic plots.*

Gesture-locked analysis. The omnibus gesture-locked analysis showed a significant Informativeness x ROI interaction effect in the time-window 200-100 ms preceding gesture initiation, $F(2,52) = 4.15$, $p = .017$, $\eta_p^2 = .15$. However, this effect did not reflect a significant main effect of Informativeness in any of the ROIs separately (all *p*'s >.14)

Speech-locked analysis. The only significant effect in the omnibus analysis locked to speech-onset was a significant Informativeness x ROI interaction effect in the time-window 500-400 ms preceding speech onset, F (2,48) = 3.18, p = .049, η_p^2 = .12. This effect reflected a trend towards a main effect of Informativeness in the anterior ROI in this time-window, F (1,23) = 3.78, p = .064, η_p^2 = .14 which was absent in other ROIs (all F's < 2.4). The anterior finding reflected a more negative ERP wave for the More Informative condition compared to the Less Informative condition (see Figure 7).

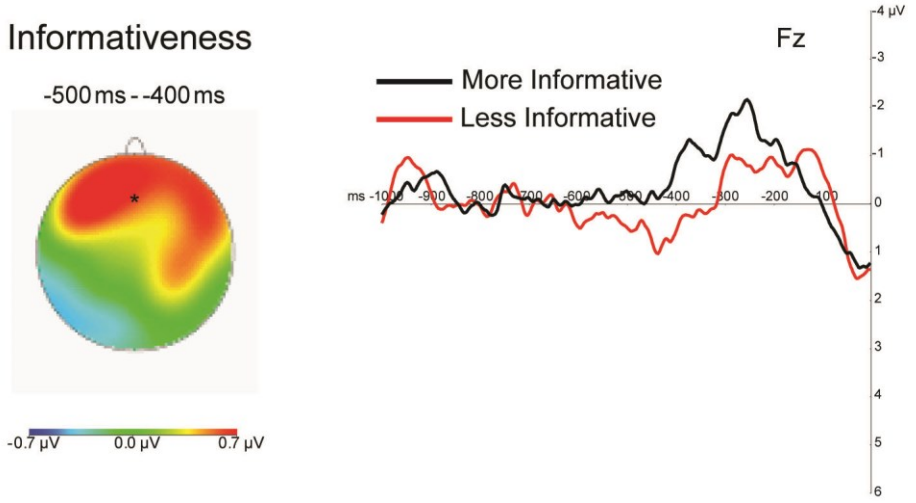

Figure 7. *Grand average waveforms and topographic plot corresponding to the voltage difference between conditions in the speech-locked ERP analysis in Experiment 2 for the main effect of Informativeness. The electrode site used for the waveforms is indicated in the corresponding topographic plot.*

Discussion

Experiment 2 revealed that the kinematic effects obtained in Experiment 1 were largely specific to situations in which gesture carried the main informational burden. Unlike in Experiment 1, in Experiment 2 no effects of Informativeness were found in the time in which apex was reached or in the duration of the post-stroke hold phase. However, a small effect of Informativeness was found in the

duration of the stroke of the gesture, with a longer stroke in case of more informative speech. Similar to Experiment 1, no effects of Informativeness were found in the speech that participants produced. In comparison with the speech-only condition, the concurrent production of a gesture delayed the onset of speech, prolonged the speech duration, and enhanced its loudness. The stimulus-locked ERP data replicated the P300 effect obtained in Experiment 1, hence suggesting that this effect is modality-independent[1]. A trend towards a frontal ERP effect of Informativeness was found preceding the onset of speech. We will discuss the theoretical implications of the findings of both experiments in the General Discussion.

General Discussion

Two experiments were carried out to further our understanding of how our intentions shape our actions in the specific case of the planning and production of pointing gestures and speech to single out a visible referent, a core everyday human communicative act (Kita, 2003; Tomasello, 2008). Specifically, we investigated whether, and if so how, the kinematics of pointing gestures are shaped by one's communicative intentions, and whether this is modulated by the presence of concurrent speech. In addition, we explored the neural and cognitive mechanisms involved in the planning of communicative pointing gestures and speech.

Behaviorally, the first experiment showed that the kinematics of a pointing gesture vary as a function of the speaker-gesturer's communicative intent. Specifically, the duration of the stroke (and as such its velocity and the moment the apex is reached) was used in order to be informative. Presumably

[1] An additional repeated measures ANOVA with the between-subject factor Experiment (2: Exp1, Exp2) and the within-subject factors ROI (2, left posterior, right posterior), and Time-Window (2: 200-300 ms, 300-400 ms) was performed on the the P3b effect (less informative average ERP *minus* more informative average ERP), calculated for each subject in both time-windows. This analysis did not show any significant main or interaction effect of Experiment, indicating that the size of the P3b effect did not differ statistically across the two experiments.

this was done in order to be as precise as possible in pointing at a target, which could be achieved by pointing more slowly. An additional benefit would then be that the addressee would have more time to identify towards which referent the gesture was heading. In addition, participants prolonged the post-stroke hold-phase of their pointing gesture presumably in order to assure that the addressee had enough time to identify which referent she pointed to. The fact that people slow down their movement in order to be more informative generalizes to instrumental actions such as reach-to-grasp movements (Becchio et al., 2012) and communicative manual actions more broadly (Vesper & Richardson, 2014). Presumably, the duration of different sub-components of the pointing gesture is not the only parameter people may use in order to communicate effectively, as previous work suggests that also the endpoint location and trajectory of a pointing gesture may be varied in relation to the location of the addressee (Cleret de Langavant et al., 2011).

In line with a previous study (Chieffi et al., 2009), in Experiment 1, the presence of speech as a second modality did not influence pointing gestures' kinematics. Other studies did find effects of the presence of speech on the kinematics of concurrently produced gestures. Gonseth et al. (2013) reported a slower gesture and a longer post-stroke hold-phase in cases where a pointing gesture was produced without speech compared to when it was produced with concurrent speech. Bernardis and Gentilucci (2006) found that participants shortened various movement phases of their symbolic gestures (e.g., a hand with protruding index-finger moving from left to right meaning 'NO') when the gesture was produced with meaningful speech compared to when it was produced in isolation. An explanation for the absence of an influence of speech production on gesture kinematics in our study is that speech was purposefully kept very simple, non-informative and repetitive across our first experiment. Increasing variation in speech (as in Gonseth et al., 2013) or adding a stronger symbolic component to it (as in Bernardis & Gentilucci, 2006) may lead to a stronger influence of speech on gesture kinematics.

The second experiment further specified that the influence of one's communicative intentions on the kinematics of one's pointing gestures is reduced in situations in which one's speech is carrying the informational burden in a multimodal referential act. For instance, participants did not use the duration of the hold phase of their gesture to be more informative in Experiment 2. Thus, when speech suffices in transmitting the required information in a certain context, one does not need to exploit the kinematics of one's gesture to the same extent as when the gesture carries the informational burden. Nevertheless, a small modulation of the duration of the gesture's stroke as a function of participants' communicative intentions was found in both experiments (i.e. a longer stroke duration to be more informative). This confirms that speech and gesture are two highly intertwined modalities in the exophoric use of referential expressions and suggests that, even when speech is carrying the most relevant information in a multimodal referential act, one's more global communicative intentions also "flow" into the gestural modality. In contrast, participants neither exploited the loudness and duration of their speech to be more informative in Experiment 2 (see Willems et al., 2010, for a similar finding). One possible explanation is that in the current task the speech content itself was informative enough such that there was no need to change any acoustic or durational parameters to be more informative.

In both our experiments, participants temporally aligned the onset of their deictic linguistic expression with the moment the pointing gesture reached its apex, regardless of whether the gesture was more or less informative. This finding is in line with previous studies showing such temporal alignment of pointing and speech (e.g., Chu & Hagoort, 2014; Levelt et al., 1985; McNeill, 1992) and with models of speech and gesture production that underline the tight synchronization of speech and gesture (e.g., De Ruiter, 2000; Krauss et al., 2000). Previous experimental studies used artificial exogenous factors (such as the application of a load to a cord attached to a participant's wrist during the execution of a gesture; Levelt et al., 1985) to investigate its effects on speech-gesture synchronization. Here, we show that also

when characteristics of the gesture vary for endogenous reasons (i.e. communicative intentions), the temporal synchrony between speech and pointing gestures is maintained.

In general, our results fit well with models of speech and gesture production that allow for a role of the speaker-gesturer's communicative intent in modulating the exact form of a gesture, such as the Sketch model (De Ruiter, 2000) and the Interface model (Kita & Özyürek, 2003). However, these models do not specify the exact sub-components of pointing gestures that people may vary on the basis of their communicative intentions. Our results suggest that duration (and as such the velocity of the stroke and the moment apex is reached) is a free parameter that people use in the execution of their pointing gestures, and further specify in which specific components (i.e. stroke or post-stroke hold) of the gesture duration is indeed varied. Even when speech is carrying the informational burden in a multimodal referential act, people's communicative intentions may lead to such use of the gesture's movement duration, as evidenced in Experiment 2. Our data cannot be explained by models of speech and gesture production that question whether the speaker's communicative intent plays a role in shaping the form of a gesture (e.g., Krauss et al., 2000).

Neurophysiologically, we observed in both experiments a stimulus-locked P3b effect preceding the production of gesture and/or speech. We argued that P3b amplitude may be modulated by task-related cognitive demands that drive attentional resource allocation, such that its amplitude is smaller when a task requires greater amounts of attentional resources (cf. Polich, 2007). Smaller amplitude of the stimulus-locked P3b in the more informative conditions in Experiment 1 may therefore reflect that participants voluntarily (Kok, 2001) allocated more attentional resources to the task when planning a more informative gesture for their addressee, independent of whether they concurrently produced speech. Experiment 2 clarified that this effect is not specific to the planning of pointing gestures, and also generalizes to situations in which referential speech is planned to describe a referent for one's addressee. This finding also confirms that the effect does not index differential visual attention paid to

the spatial location or physical properties (e.g. color) of a referent, but rather in our study reflects the allocation of domain-general attentional resources that may be used to successfully plan an action on the basis of one's (communicative) intentions.

The gesture-locked frontal ERP marker of communicative intent directly preceding the onset of the pointing gesture was specific to the case where the gesture carried the informational burden (i.e. Experiment 1). The locus of this effect modulating the readiness potential is reminiscent of a previous study investigating pointing by infants, which also identified a fronto-central marker of communicative intent measured using EEG (Henderson et al., 2002). Several other studies have also linked frontal effects in ERPs to 'mentalizing' or theory-of-mind related activations (e.g., Liu et al., 2004; Sabbagh, 2004) and recent neuroimaging studies relate activity in neuronal structures in frontal cortex to the mentalizing involved in the production and comprehension of communicative pointing (e.g., Brunetti et al., 2014). The fact that our effect reflects a modulation of the readiness potential (Kornhuber & Deecke, 1965) over fronto-central areas suggests an interaction between planning the execution of a motor program and activation of the mentalizing network (Amodio & Frith, 2006; Van Overwalle & Baetens, 2009; Willems et al., 2010). In sum, these findings underline that both intentional and modality-independent attentional mechanisms are active when one plans the execution of a communicative, referential pointing gesture for an addressee.

Finally, a speech-locked trend towards an effect of participants' communicative intentions was found 500 - 400 ms preceding the onset of their speech. Interestingly, it had an opposite directionality compared to the frontal gesture-locked effect of participants' communicative intentions. Future research is needed to verify whether the speech finding is robust. Note that, on the basis of models of speech production, the timing of the effect is where one would expect an influence of one's intentions in the speech production process (e.g., Indefrey & Levelt, 2004). The current study shows that it is worthwhile

and feasible to investigate the intentions behind speech (and gesture) production, a crucial component of the speech production process.

To conclude, we have shown that people shape the exact kinematics of their pointing gesture as a function of their specific communicative intentions, in tight temporal alignment with their speech, and particularly when the gestural modality carries the informational burden. Furthermore we have shown that both intentional and modality-independent attentional neural mechanisms are active in planning the execution of a communicative pointing gesture. These findings contribute to a better understanding of the complex interplay between action, attention, intention, and language in the core human communicative act of planning and producing referential utterances using speech and gesture.

References

Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268-277.

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*(6), 415-419.

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, 205-226.

Becchio, C., Manera, V., Sartori, L., Cavallo, A., & Castiello, U. (2012). Grasping intentions: from thought experiments to empirical evidence. *Frontiers in human neuroscience*, *6: 117*.

Bernardis, P., & Gentilucci, M. (2006). Speech and gesture share the same communication system. *Neuropsychologia*, *44*(2), 178-190.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1. 05)[Computer program].

Brunetti, M., Zappasodi, F., Marzetti, L., Perrucci, M. G., Cirillo, S., Romani, G. L., Pizzella, V., & Aureli, T. (2014). Do you know what I mean? Brain oscillations and the understanding of communicative intentions. *Frontiers in Human Neuroscience*, *8*, 36.

Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet* (pp. 9-33). Hillsdale, NJ: Erlbaum.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics*, *47*(1), 14-27.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 255, Vol. 63,* 1-174.

Chieffi, S., Secchi, C., & Gentilucci, M. (2009). Deictic word and gesture production: Their interaction.

Behavioural brain research, *203*(2), 200-206.

Chu, M., & Hagoort, P. (2014). Synchronization of speech and gesture: Evidence for interaction in action. Journal of Experimental Psychology: General. Advance online publication. doi:10.1037/a0036281

Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet* (pp. 243-268). Hillsdale NJ: Erlbaum.

Cleret de Langavant, L., Remy, P., Trinkler, I., McIntyre, J., Dupoux, E., Berthoz, A., & Bachoud-Lévi, A. C. (2011). Behavioral and neural correlates of communication via pointing. *PloS one*, *6*(3), e17719.

Cooperrider, K. (2011). Reference in action: Links between pointing and language. Doctoral dissertation, University of California, San Diego.

Cooperrider, K., & Núñez, R. (2012). Nose-pointing: Notes on a facial gesture of Papua New Guinea. *Gesture*, *12*(2), 103-129.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141-168.

De Ruiter, J. P. (1998). *Gesture and speech production.* Doctoral dissertation, University of Nijmegen, The Netherlands.

De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and Gesture* (pp. 284-311). Cambridge: Cambridge University Press.

Enfield, N. J., Kita, S., & De Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, *39*(10), 1722-1741.

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture*, *4*(2), 157-195.

Gonseth, C., Vilain, A., & Vilain, C. (2013). An experimental study of speech/gesture interactions and distance encoding. *Speech communication*, *55*(4), 553-571.

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*(2), 95-112.

Henderson, L. M., Yoder, P. J., Yale, M. E., & McDuffie, A. (2002). Getting the point: Electrophysiological correlates of protodeclarative pointing. *International Journal of Developmental Neuroscience*, *20*(3), 449-458.

Holler, J., & Stevens, R. (2007). The effect of common ground on how speakers use gesture and speech to represent size information. *Journal of Language and Social Psychology*, *26*(1), 4-27.

Holler, J., & Wilkin, K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics*, *43*(14), 3522-3536.

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1), 101-144.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological science*, *16*(5), 367-371.

Kendon, A. (1988). *Sign languages of Aboriginal Australia: Cultural, semiotic and communicative perspectives*. Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Kita, S. (2003). *Pointing. Where language, culture, and cognition meet.* Hillsdale, NJ: Erlbaum.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and language*, *48*(1), 16-32.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*(3), 557-577.

Kornhuber, H. H., & Deecke, L. (1965). Hirnpotentialänderungen bei Willkürbewegungen und passiven

Bewegungen des Menschen: Bereitschaftspotential und reafferente Potentiale. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere*, *284*(1), 1-17.

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261-283). Cambridge: Cambridge University Press.

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation*, *11*(2), 417-441.

Levelt, W. J., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language*, *24*(2), 133-164.

Liu, D., Sabbagh, M. A., Gehring, W. J., & Wellman, H. M. (2004). Decoupling beliefs from reality in the brain: an ERP study of theory of mind. *NeuroReport*, *15*(6), 991-995.

McNeill, D. (1992). *Hand and Mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

Melinger, A., & Levelt, W. J. (2004). Gesture and the communicative intention of the speaker. *Gesture*, *4*(2), 119-141.

Moore, C., & D'Entremont, B. (2001). Developmental changes in pointing as a function of attentional focus. *Journal of Cognition and Development*, *2*(2), 109-129.

Mundy, P., Card, J., & Fox, N. (2000). EEG correlates of the development of infant joint attention skills. *Developmental psychobiology*, *36*(4), 325.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of

addressee location on representational gestures. *Journal of Memory and Language*, *46*(4), 688-704.

Petrarch, F. (1898/1336). The ascent of Mount Ventoux. In J. H. Robinson (Ed.), *Petrarch: The first modern scholar and man of letters* (pp. 307-320). New York: Putnam.

Pierno, A. C., Tubaldi, F., Turella, L., Grossi, P., Barachino, L., Gallo, P., & Castiello, U. (2009). Neurofunctional modulation of brain regions by the observation of pointing and grasping actions. *Cerebral Cortex*, *19*(2), 367-374.

Polich, J. (2007). Updating P300: an integrative theory of P3a and P3b. *Clinical neurophysiology*, *118*(10), 2128-2148.

Sabbagh, M. A. (2004). Understanding orbitofrontal contributions to theory-of-mind reasoning: Implications for autism. *Brain and cognition*, *55*(1), 209-219.

Sartori, L., Becchio, C., Bara, B. G., & Castiello, U. (2009). Does the intention to communicate affect action kinematics?. *Consciousness and cognition*, *18*(3), 766-772.

Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: the role of movement information. *Cognition*, *119*(2), 242-252.

Sherzer, J. (1973). Verbal and nonverbal deixis: The pointed lip gesture among the San Blas Cuna. *Language in Society*, *2*(01), 117-131.

Southgate, V., Van Maanen, C., & Csibra, G. (2007). Infant pointing: Communication to cooperate or communication to learn?. *Child development*, *78*(3), 735-740.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78,* 705-722.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, *48*(3), 564-584.

Vesper, C. & Richardson, M. (2014). Strategic communication and behavioral coupling in asymmetric

joint action. *Experimental Brain Research, 232*(9), 2945-2956.

Willems, R. M., de Boer, M., de Ruiter, J. P., Noordzij, M. L., Hagoort, P., & Toni, I. (2010). A dissociation between linguistic and communicative abilities in the human brain. *Psychological Science*, *21*(1), 8-14.

Wilkins, D. (2003). Why pointing with the index finger is not a universal (in sociocultural and semiotic terms). In S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet* (pp. 171-215). Hillsdale NJ: Erlbaum.

Chapter 5

The Neural Integration of Pointing Gestures and Speech in a Visual Context: An fMRI Study

Based on: Peeters, D., Snijders, T. M., Hagoort, P., & Özyürek, A. (under review). The neural integration of pointing gestures and speech in a visual context: An fMRI study.

Abstract

Comprehension of pointing gestures is ontogenetically and phylogenetically fundamental to human communication. However, the neural mechanisms that subserve the integration of pointing gestures and speech in a visual context in comprehension are unclear. Here we present the results of an fMRI study in which participants watched static images of an actor pointing at an object while they listened to her referential speech. Both a mismatch paradigm and a bimodal enhancement manipulation were employed. It was found that bilateral auditory and visual areas and left premotor regions were involved in the concomitant perception of speech, gesture and referent, whereas left inferior frontal gyrus subserved the semantic unification of referential gesture and speech in a triadic context. These findings suggest an important role for primary areas in audiovisual binding and confirm the importance of left inferior frontal gyrus in semantically integrating information from multiple modalities across semiotic domains.

The Neural Integration of Pointing Gestures and Speech in a Visual Context: An fMRI Study

> "While all gestures cry out for attention,
> pointing gestures immediately deflect
> that attention elsewhere" (Cooperrider, 2011, p.7)

Pointing gestures are a fundamental part of human communication (Kita, 2003). By producing them in everyday life we connect our communication to entities in the world around us (Clark, 2003). In establishing a triadic link between child, caregiver, and referent, they play a crucial role in language acquisition (Butterworth, 2003; Carpenter, Nagell, & Tomasello, 1998; Iverson & Goldin-Meadow, 2005; Tomasello, Carpenter, & Liszkowski, 2007) and impairments in the production and comprehension of pointing gestures are an early marker of the neurodevelopmental disorder autism (e.g. Baron-Cohen, 1989). From a phylogenetic viewpoint, it has been claimed that (declarative) pointing is a uniquely human form of communication in a natural environment (Call & Tomasello, 1994; Kita, 2003; Tomasello et al., 2007).

Previous neuroimaging work investigating the comprehension of index-finger pointing gestures has presented the gestures in a context that lacked both a larger visual triadic context and co-occurring speech (e.g. Materna, Dicke, & Thier, 2008; Sato, Kochiyama, Uono, & Yoshikawa, 2009). However, in everyday human referential communication pointing gestures often occur in a context in which one perceives not only the person pointing but also the referent she points at and the speech she may concomitantly produce. It is currently unclear how in such situations input from different modalities (visual: speaker, pointing gesture, referent; auditory: speech) is integrated in the brain. The lack of empirical neurocognitive research in this domain is surprising, because comprehending and integrating our interlocutors' referential (i.e. deictic) gesture and speech in a visual context is often critical to understand what they are talking about and a core feature of everyday communication (Bühler, 1934; Clark, 1996; Kendon, 2004). The

current study therefore investigates the neural mechanisms underlying the audiovisual (multimodal) and semantic integration of manual pointing gestures with speech in a visual, triadic context.

The majority of studies investigating the neural integration of gestures with co-occurring speech have focused on *iconic* co-speech gestures, i.e. hand movements that visually resemble the meaning of the linguistic part of the utterance they accompany (Kendon, 2004; McNeill, 1992), such as when moving up and down one's hand while talking about a basketball game. Previous work has shown that recipients semantically process these everyday communicative actions (e.g., Kita & Özyürek, 2003; Özyürek, Willems, Kita, & Hagoort, 2007). Critically, previous studies investigating the neural integration of such iconic gestures with speech differ in the functional roles they attribute to i) the (posterior portions of the) superior temporal sulcus (STS), the superior temporal gyrus (STG) and the middle temporal gyrus (MTG), and ii) the left inferior frontal gyrus (LIFG) in the integration of audiovisual and semantic information from gesture and speech in comprehension (for overviews, see Andric & Small, 2012; Dick, Mok, Beharelle, Goldin-Meadow, & Small, 2014; Marstaller & Burianová, 2014; Özyürek, 2014).

It is relatively uncontroversial that LIFG, more specifically its pars triangularis, plays a role in the integration of speech and gesture (e.g. Dick et al., 2014; Skipper, Goldin-Meadow, Nusbaum, & Small, 2007; Willems, Özyürek, & Hagoort, 2007; 2009; but see Holle, Obleser, Rueschemeyer, & Gunter, 2010), possibly in interplay with MTG (Dick et al., 2014; Green, Straube, Weis, Jansen, Willmes, et al., 2009). Willems et al. (2007) were the first to study the integration of speech and gesture using fMRI. In an orthogonal design, the ease of integration of linguistic and gestural information into a preceding sentence context was manipulated. An increase in activation in LIFG was found when words and/or gestures were incongruent

("mismatch conditions") compared to when they were congruent ("match condition") with preceding speech (see also Özyürek et al., 2007). Such findings confirm LIFGs status as a multimodal integration site that plays a crucial role in the semantic unification of information from different modalities (Hagoort, 2005; 2013). Such accounts argue, however, that LIFG is a node in a larger network that subserves the integration of gesture and speech, and also attribute a role to STS/STG and MTG in the perception and integration of speech-gesture combinations (e.g. Dick et al., 2014; Willems et al., 2009).

However, the specific functional role of the posterior part of the STS and the adjacent STG and MTG in the larger speech-gesture integration process is unclear. Holle et al. (2008; 2010) argue that left pSTS subserves the integration of gesture and speech at a semantic (conceptual) level. Holle et al. (2008) presented participants with videos in which a speaker uttered sentences that contained an ambiguous word (e.g. *He looked at the pass*) that was combined with either a gesture related to the more frequent meaning of the word, a gesture related to its less frequent meaning, or a grooming (control) movement. The authors report enhanced activity in left posterior STS for co-speech gestures compared to grooming movements and conclude that this activation reflects "multimodal semantic interaction between a gesture and its co-expressive speech unit" (Holle et al., 2008, p. 2022). In sharp contrast, Dick et al. (2014) recently argued that pSTS is not involved in *semantic* integration per se, but rather in connecting information from visual and auditory modalities in general. Willems et al. (2009) argue that pSTS/MTG shows enhanced activation when different sources of information converge on a common memory representation, as in the case of perceiving pantomimes and concurrent, matching speech (see also Hagoort, Baggio, & Willems, 2009), in contrast with higher-order semantic unification subserved by LIFG.

As outlined above, in the current study we focus on a different type of gesture, namely (deictic) pointing gestures. Unlike iconic gestures, pointing gestures in exophoric use canonically create a vector towards a referent to shift the gaze of an addressee and establish a joint focus of attention (Kita, 2003). Furthermore, whereas speech and iconic gestures often allow communicating about entities that are not immediately physically present ("displacement", Hockett, 1960; Perniss & Vigliocco, 2014), pointing gestures in exophoric use play a crucial role in referential communication about entities that speaker and addressee may perceive in the immediate extra-linguistic context of a conversation. Therefore, the integration of speech and pointing gestures towards a referent need not necessarily recruit the same neural and cognitive mechanisms as in the integration of speech with iconic or other type of gestures (e.g., Hubbard, Wilson, Callan, & Dapretto, 2009; Quandt, Marshall, Shipley, Beilock, & Goldin-Meadow, 2012).

Although it is currently unknown which cortical areas are involved in integrating pointing gestures and speech, a number of studies have looked at the neural correlates of comprehending pointing gestures in isolation and at their integration with other cues such as the gesturer's gaze direction (e.g., Brunetti, Zappasodi, Marzetti, Perrucci, Cirillo, et al., 2014; Conty, Dezecache, Hugueville, & Grèzes, 2012; Gredebäck, Melinder, & Daum, 2010; Materna et al., 2008; Sato et al., 2009). Sato et al. (2009), for instance, showed that the perception of a (meaningless) pointing hand, compared to a non-directional closed hand, elicits enhanced activation in a network of mainly right-hemisphere regions, including right IFG, right angular gyrus, right parietal lobule, right thalamus, and bilateral lingual gyri. Materna et al. (2008) suggest that bilateral posterior STS is involved in following the direction of a pointing finger. Conty et al. (2012) show that integration of pointing gestures and gaze direction in comprehension recruits parietal and

supplementary motor cortices in the right hemisphere. All in all, these findings suggest an extensive right-hemisphere dominant network that is activated when one perceives a manual pointing gesture that shifts one's attention.

Finally, Pierno, Tubaldi, Turella, Grossi, Barachino, et al. (2009) compared the observation of a static image of a hand pointing at an object to the observation of a hand grasping an object and to a control condition of a hand resting next to an object. Compared to the control condition, the perception of the pointing hand and object elicited enhanced activation in left MTG, left parietal areas (postcentral gyrus and supramarginal gyrus) and left middle occipital gyrus. However, the pointing condition did not recruit significant differential activity compared to the grasping condition. Nevertheless these results suggest that, in addition to the right-lateralized network involved in perceiving a pointing hand, a left-lateralized set of cortical areas may be involved in visually integrating a pointing hand and an object.

The present study

In the present study, we investigated which cortical regions subserve the integration of pointing gestures with speech in a visual, everyday context. In an event-related functional magnetic resonance imaging (fMRI) study, participants were presented with images of a speaker who pointed at one of two different objects as they listened to her speech. We employed a mismatch paradigm, such that speech either referred to the object the speaker pointed at or to the other visible object. As such, speech and gesture were individually always correct, but there was congruence or incongruence when semantically integrated in the larger visual context. Thus, the mismatch-match comparison taps into the semantic integration/unification of pointing gestures and speech. In addition we included two unimodal runs (audio-only and visual-only; cf. Willems

et al., 2009) to test for possible bimodal enhancement (audiovisual > audio-only + visual-only). Both mismatch paradigms and bimodal-unimodal comparisons have been successfully used in the past to study the integration of iconic gestures and speech (e.g. Dick et al., 2014; Willems et al., 2009) and audiovisual integration more broadly (e.g., Belardinelli, Sestieri, Di Matteo, Delogu, Del Gratta, et al., 2004; Calvert, Campbell, & Brammer, 2000).

Because this is the first study investigating the neuronal integration of pointing gestures with speech in comprehension, predictions were derived on the basis of previous speech-gesture integration studies that used *iconic* gestures in their stimulus materials. If LIFG plays a key role in the semantic integration of gesture and speech (Dick et al., 2014; Skipper et al., 2007; Willems et al., 2009), it should show enhanced activation in the mismatch compared to the match condition. This is in line with a view of LIFG as a modality-independent multimodal integration site, with its pars triangularis specifically involved in semantic unification of information from different input streams (e.g. Hagoort, 2013; Willems et al., 2007).

Furthermore, if pSTS is mainly involved in "connecting information from the visual and auditory modalities" (Dick et al. 2014, p. 914), it should be sensitive to the bimodal enhancement comparison. Similarly, if pSTS/MTG is involved in mapping different sources of information onto a common memory representation (Hagoort et al., 2009; Willems et al., 2009), then it should also show enhanced activation in the bimodal (match) condition compared to the sum of unimodal conditions, because in the bimodal condition speech and object match whereas in the unimodal conditions speech and gesture are presented in isolation. Conversely, if multimodal semantic integration of gesture and speech recruits the posterior part of the STS region (Holle et al., 2008; 2010), then this region should show enhanced activation in the mismatch-match comparison.

Finally, we included two conditions in which one of the two objects in the images was highlighted by an attentional cue in the absence of gesture. This allowed investigating whether the possible role of LIFG in semantic unification of speech and pointing gesture in a triadic context was dependent on the perceived communicative intentions of the gesturer. Research by Kelly and colleagues suggests that speech-gesture integration differs from the integration of gestures with actions more broadly because the former are generally viewed as more intended to accompany the speech signal compared to the latter (e.g., Kelly, Healey, Özyürek, & Holler, 2014). Pointing gestures are shaped by the communicative intentions of the gesturer (Chapter 4 of this thesis), and in that sense differ from other cues in the environment that may shift our attention. Therefore the integration of pointing gestures with speech may differ from the integration of other attentional cues with concurrently perceived speech. In sum, the current study thus aims to shed more light on the functional roles of different cortical areas involved in speech-gesture integration by investigating the integration of speech with a novel type of gesture, namely index-finger pointing.

Method

Participants

Twenty-three right-handed (Oldfield, 1971) native speakers of Dutch (18 female; mean age 23.6, range 18-29) participated in the experiment. Data from three additional participants were discarded due to technical failure (n = 2) or drowsiness (n = 1). Participants had normal or corrected-to-normal vision, no language or hearing impairments or history of neurological disease. They provided written informed consent and were paid for participation.

Stimulus Materials and Experimental Design

The experimental materials consisted of 40 spoken items in Dutch of the form "definite article + noun" (e.g., "het kopje", *the cup*), 80 pictures in which a model (henceforth: the speaker) pointed (index-finger extended; Kendon, 2004) at one of two objects presented at a table in front of her (henceforth "target pictures"), and 80 pictures that were the same except that one of the two objects was framed by a green box and that the speaker did not point (henceforth "attentional pictures"). The 40 spoken items were spoken at a normal rate by a female native speaker of Dutch, recorded in a sound proof booth, and digitized at a sample frequency of 44.1 kHz. They were equalized in maximum amplitude using *Praat* software (version 5.2.46; Boersma & Weenink, 2009) and had an average duration of 837 ms (*SD* = 155 ms). In half of the target pictures the speaker pointed at the object at her left and in the other half of the target pictures she pointed at the object at her right. Similarly, in half of the attentional pictures the object at her left was framed and in the other half the object at her right. The 40 different table-top objects in the pictures were selected on the basis of a pre-test reported elsewhere (Chapter 3 of this thesis) that confirmed that these objects elicited highly consistent labels (i.e. > 90% naming consistency for each object across 16 participants) across individuals from the same participant pool as the current participants.

The experiment consisted of three blocks. The *speech-only* block (AUDIO) consisted of the 40 spoken items. The *picture-only* block (VISUAL) consisted of 40 pictures in which the speaker pointed at an object. The *mixed block* consisted of 160 speech-picture pairs that made up four conditions. In the Bimodal Match (BM) condition, the spoken stimulus matched the object the speaker pointed at. In the Bimodal Mismatch (BMM) condition, the spoken stimulus did not match the object she pointed at but the other object. In the Attentional Match (AM) condition,

the spoken stimulus matched the framed object. In the Attentional Mismatch (AMM) condition, the spoken stimulus matched the object that was not framed. Each condition consisted of 40 speech-picture pairs. Figure 1 shows a subset of pictures used in the experiment.

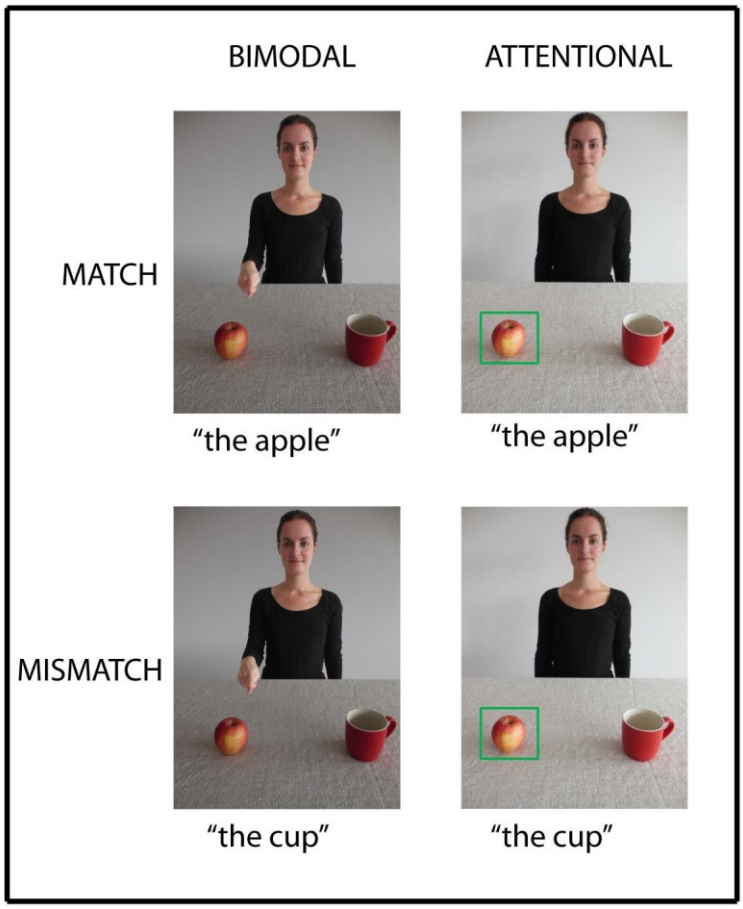

Figure 1. *Overview of the four conditions involved in the mismatch manipulation.*

Procedure

The three blocks were presented sequentially with specific instructions preceding each block. The order of presentation of the blocks was counterbalanced across participants. All stimuli were presented in an event-related design and in a randomized order. Twelve different randomized lists were used. The *speech-only block* consisted of the presentation of the 40 spoken

stimuli. A trial in this block consisted of a fixation cross presented for a jittered duration of 2-6s followed by the presentation of the spoken stimulus. The *picture-only block* consisted of the presentation of 40 pictures in which the speaker pointed at one of the two objects. No speech was presented during this block. A trial in this block consisted of a fixation cross presented for a jittered duration of 2-6s followed by the presentation of the picture for 2s. The *mixed block* consisted of 160 target trials in which a fixation cross (jittered duration of 2-6s) was followed by the presentation of a picture (for 2s) with a concurrently presented spoken stimulus. The onset of the spoken stimulus was 50 ms after the onset of the picture presentation. In both the picture-only block and the mixed block, the speaker pointed at the object at her left in half of the cases, and at the object at her right in the other half of the cases. In the mixed block, in half of the attentional pictures the object at the speaker's left was framed and in the other half of the attentional pictures the object at her right.

Pictures were presented on the screen using *Presentation* software (Neurobehavioral Systems) and speech was presented through nonmagnetic headphones that reduced scanner noise. Participants looked at the screen via a mirror mounted to the head coil. The size of the pictures on the screen was determined on the basis of judgments from two pilot subjects that did not participate in the main experiment. They confirmed that all objects, the speaker's gesture, and the attentional markers, were clearly visible while focusing on the center of the screen.

Participants in the main experiment were instructed to carefully listen to the speech and look at the pictures. They were asked to press a button with the middle finger of their left hand when an item (i.e. a spoken stimulus in the speech-only block, a picture in the picture-only block, and the picture-speech pair in the mixed block) was exactly the same on two subsequent trials. In the speech-only block and the picture-only block, four stimuli were repeated on two subsequent

trials. In the mixed block 16 stimuli were repeated on two subsequent trials. The second presentations of such items thus served as catch trials eliciting a button press and were excluded from further MRI analyses.

The experiment was preceded by a practice session that consisted of three blocks of nine trials each (i.e. eight items of which one was repeated and served as a catch trial to familiarize participants with the task). Before the start of the practice block the scanner was switched on and a number of spoken stimuli were played in order to adjust the volume level of the spoken items. Participants were asked to indicate whether the volume should go up or down. The items used in this audio check and the items used in the practice blocks were not used in the main experiment.

fMRI Data Acquisition

Participants were scanned with a Siemens 3-T Skyra MRI scanner using a 32-channel head coil. The functional data were acquired in one run using a multiecho echo-planar imaging sequence (Poser, Versluis, Hoogduin, & Norris, 2006), in which image acquisition happens at multiple echo times (TEs) following a single excitation [time repetition (TR) = 2250 ms; TE1 = 9 ms; TE2 = 19.5 ms; TE3 = 30 ms; TE4 = 40 ms; echo spacing = 0.51 ms; flip angle = 90 °]. This procedure broadens T2* coverage and improves T2* estimation (see Poser et al., 2006, for details). Each volume consisted of 36 slices of 3 mm thickness [ascending slice acquisition; voxel size = 3.3 x 3.3 x 3 mm; slice gap = 10 %; field of view (FOV) = 212 mm]. The first 30 volumes preceded the start of the presentation of the first stimulus and were used for weight calculation of each of the four echoes. Subsequently, the 31st volume was taken as the first volume in preprocessing. The functional run was followed by a whole-brain anatomical scan using a high resolution *T*1-weighted magnetization-prepared, rapid gradient echo sequence

(MPRAGE) consisting of 192 sagittal slices (TR = 2300 ms; TE = 3.03 ms; FOV = 256 mm; voxel size = 1 x 1 x 1 mm) accelerated with GRAPPA parallel imaging.

Data Analysis

Data were analyzed using statistical parametric mapping (SPM8; www.fil.ion.ucl.ac.uk/spm/) implemented in Matlab (Mathworks Inc., Sherborn, MA, USA). The four echoes of each volume were combined to yield one volume per TR (Poser et al., 2006), after which standard pre-processing was performed [realignment to the first volume, slice acquisition time correction to time of acquisition of the middle slice, coregistration to T1 anatomical reference image, normalization to Montreal Neurological Institute (MNI) space (EPI template), smoothing with an 8 mm full-width at half-maximum (FWHM) Gaussian kernel, and high-pass filtering (time-constant = 128 s)] (Friston, Holmes, Poline, Grasby, Williams, et al., 1995).

Statistical analysis was performed in the context of the general linear model (GLM). Stimulus onset (i.e. the onset of the picture in all conditions, except the speech-only condition in which it was the onset of speech) was modeled as the event of interest for each condition. Each condition thus contained 40 events. The 6 condition regression parameters were convolved with a canonical hemodynamic response function. Additionally, 6 motion parameters from the realignment preprocessing step were included in the first-level model.

A whole-brain analysis was performed by entering first-level contrast images of each of the six conditions > baseline for each participant into a flexible factorial model at second-level (with factors Condition [6] and Participant [23]). To test for possible bimodal enhancement, two analyses were performed. First, the bimodal match condition was compared to the sum of the unimodal conditions (BM > AUDIO + VISUAL). Contrast weights were balanced such that the

bimodal condition was weighted twice as strongly as each unimodal condition. This comparison indicates whether any areas are activated in the bimodal condition in addition to areas activated in the unimodal conditions (unimodal areas will also appear in this comparison as the bimodal condition is weighted twice). Second, a conjunction analysis, testing a logical AND (Nichols, Brett, Andersson, Wager, & Poline, 2005), was performed to subsequently verify whether any unimodal areas were activated more in the bimodal compared to the unimodal presentation of the stimuli. This analysis was implemented as (BM > AUDIO ∩ BM > VISUAL), inclusively masked with the conjunction of the unimodal conditions compared to zero, thus yielding the comparison 0 < AUDIO < BM > VISUAL > 0. Furthermore, two analyses were performed to compare audiovisual mismatch to audiovisual congruency. First, the bimodal mismatch condition was compared to the bimodal match condition (BMM > BM). Second, the attentional mismatch condition was compared to the attentional match condition (AMM > AM).

Whole-brain correction for multiple comparisons was applied by combining a significance level of $p = 0.001$ (uncorrected at the voxel level) with a cluster extent threshold using the theory of Gaussian random fields (Friston, Holmes, Poline, Price, & Frith, 1996). All clusters are reported at an alpha level of $p < 0.05$ family-wise error (FWE) corrected across the whole brain (Hayasaka & Nichols, 2003).

We had the a priori hypothesis that LIFG would be recruited more in the BMM condition compared to the BM condition as this comparison arguably taps into semantic integration/unification of speech and gesture. However, it is unclear whether such a potential involvement of LIFG is specific to communicatively intended gestures and speech or, instead, generalizes to any semantic speech-referent relation as induced by an attentional cue (i.e. it would also show up in the AMM-AM comparison). Therefore, a region-of-interest (ROI)

analysis was performed in LIFG. The ROI was an 8 mm sphere around centre voxels in LIFG taken from a meta-analysis on a large number of neuroimaging studies of semantic processing (Vigneau, Beaucousin, Herve, Duffau, Crivello, et al., 2006; cf. Willems et al., 2009). MNI coordinates were [-42 19 14]. Contrast estimates were calculated for each participant at first-level for the four conditions (AM, AMM, BM, BMM) using Marsbar (http://marsbar.sourceforge.net/).

Results

Behavioral performance

Participants detected 91.5 % of all catch trials. These data were not further analyzed.

Whole-brain analysis

We first tested for possible bimodal enhancement by comparing the bimodal congruent condition to the sum of the unimodal conditions: (BM > AUDIO + VISUAL). This analysis revealed increased activations in a network of areas including bilateral occipital areas, right STG and left STG/MTG, and left premotor areas (see Table 1 and Fig. 2). The reverse contrast (AUDIO + VISUAL > BM) did not yield any significant cluster that survived statistical threshold. Table 1 and Figure 2 present the results of this analysis, as well as the comparisons of the unimodal AUDIO and VISUAL conditions to baseline. The conjunction analysis (0 < AUDIO < BM > VISUAL > 0) failed to show any cluster that survived the statistical threshold (no voxels <.001 uncorrected).

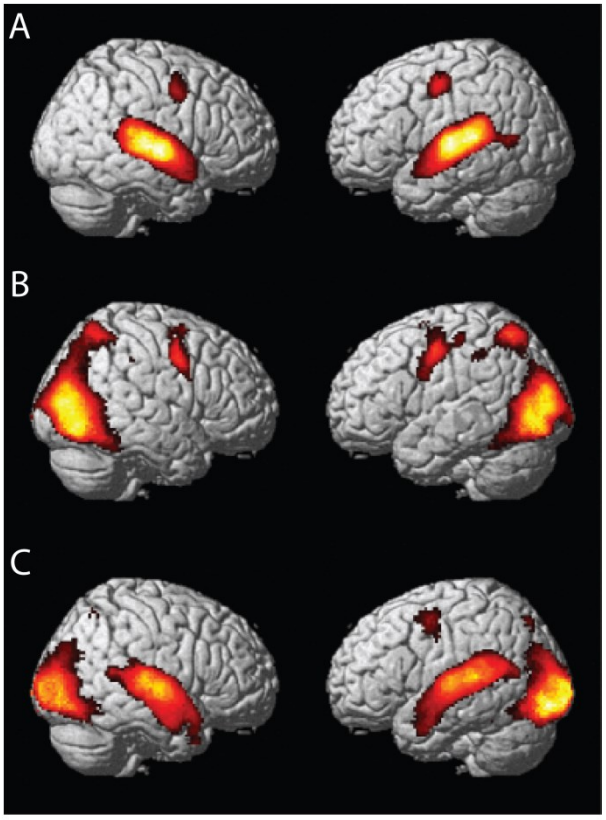

Figure 2. *Results from the whole brain analysis comparing (A) AUDIO > 0, (B) VISUAL > 0, and (C) BM > AUDIO + VISUAL. Results are displayed at p < .05, family-wise error corrected at the cluster-level.*

Table 1. *Results of the whole-brain analyses comparing unimodal to bimodal conditions.* p-*values are at the cluster-level, FWE-corrected.*

Contrast	*p*	k (extent)	*t*-value	MNI coordinates			Region/Peak
A > 0	.000	4665	13.19	-58	-16	0	left middle/superior temporal gyrus
			13.08	-52	-18	6	
			12.90	-60	-30	10	
	.000	4088	15.96	60	-12	0	right superior temporal gyrus
			13.78	64	-20	4	
			12.68	56	-24	6	
	.000	381	6.48	-52	-10	50	left postcentral gyrus
	.007	238	5.24	52	0	48	right precentral gyrus
V > 0	.000	7735	11.80	46	-80	2	right inferior/middle occipital gyrus, right inf. temporal gyrus
			11.67	38	-74	-12	
			11.36	42	-54	-14	
	.000	6878	12.64	-42	-76	-2	left middle occipital gyrus, left fusiform gyrus,
			10.45	-32	-52	-16	
			10.11	-32	-70	-12	
	.000	1187	6.75	-8	6	54	left supplementary motor area, right precentral gyrus
			6.36	52	2	38	
			5.14	36	-4	50	
	.000	1165	6.10	-30	-8	50	left precentral gyrus
			6.08	-44	-8	44	
			5.46	-24	-12	54	
BM > A + V	.000	12861	11.29	-14	-96	-8	left inferior occipital gyrus, left fusiform gyrus
			11.08	-28	-84	-8	
			10.45	-20	-88	-10	
	.000	5116	9.94	60	-12	0	right superior temporal gyrus
			9.20	52	-20	6	
			7.34	54	0	-8	
	.000	5068	8.22	-52	-16	6	left middle/superior temporal gyrus
			8.19	-58	-32	8	
			8.15	-50	-36	10	

Table 1. *(continued).*

Contrast	*p*	k (extent)	*t*-value	MNI coordinates			Region/Peak
	.000	476	5.34	-22	-28	-4	left/right thalamus
			4.99	20	-28	-2	
			4.56	26	-24	-6	
	.008	230	4.98	-52	-4	52	left precentral/postcentral gyrus
			3.93	-42	2	56	
			3.75	-54	-8	44	
0 < A < BM > V > 0		-		-	-	-	

Abbreviations: A, Audio; V, Visual; BM, Bimodal Match

Second, we compared the mismatch conditions to the match conditions. Contrasting BMM with BM showed increased activations in left inferior frontal gyrus (Fig. 3 and Table 2). The reverse contrast (BM > BMM) did not show any significant cluster that survived the statistical threshold. Also contrasting AMM with AM did not show any areas that survived the statistical threshold (Table 2).

Figure 3. *Results from the whole brain analysis comparing Bimodal Mismatch (BMM) > Bimodal Match (BM). Results are displayed at p < .05, family-wise error corrected at the cluster-level.*

ROI analysis

An ROI analysis was performed comparing mismatch to match conditions in the predefined ROI (8 mm sphere around MNI coordinates -42 19 14) in LIFG. The interaction between cue (pointing gesture / attentional cue) and congruency (match / mismatch) failed to reach significance, F (1,22) = 2.10, p = .162. However, dependent samples t-tests revealed that there was enhanced activation in LIFG in mismatch vs. match conditions when the speaker's pointing gesture indicated the referent object, t(22) = -2.43, p = .024. There was no difference in activation in the ROI between the attentional mismatch and match conditions, t(22) = .48, p = .637. Figure 4 presents the contrast estimates for the four conditions.

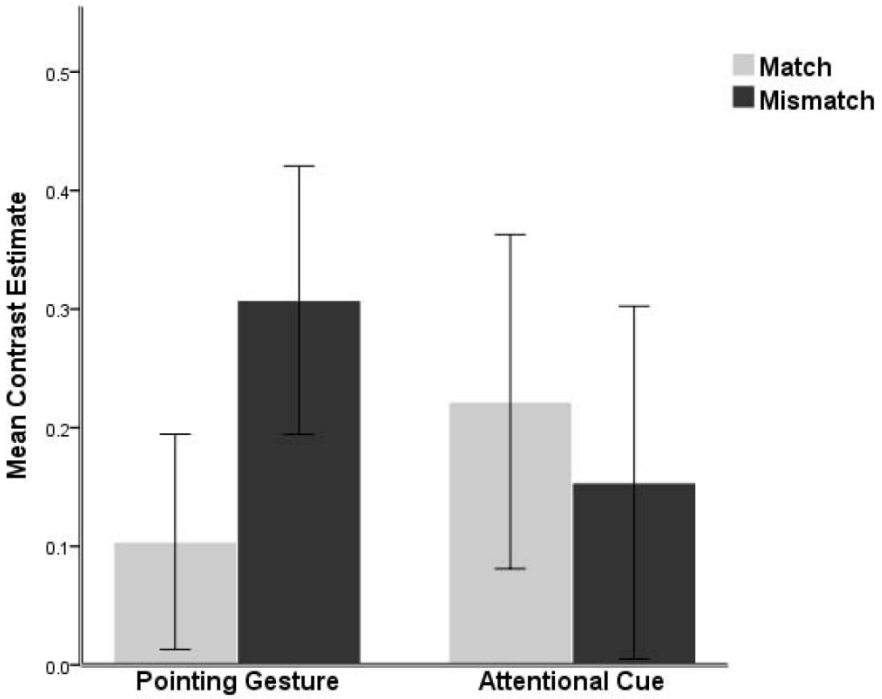

Figure 4. *ROI results. Mean contrast estimates for AM, AMM, BM, and BMM. Error bars represent standard errors around the mean.*

Table 2. *Results of the whole-brain analyses comparing congruent (match) to incongruent (mismatch) conditions.* p-*values are at the cluster-level, FWE-corrected.*

Contrast	*p*	k (extent)	*t*-value	MNI coordinates			Region/Peak
BMM - BM	.01	220	4.01	-46	20	20	Left inferior frontal gyrus (pars triangularis)
			3.72	-36	18	20	
			3.69	-50	28	18	
AMM - AM		-		-	-	-	

Abbreviations: AM, Attentional Match; AMM, Attentional Mismatch; BM, Bimodal Match; BMM, Bimodal Mismatc

Discussion

The present study investigated the neural integration of pointing gestures and speech in a visual, triadic context in comprehension. Both a mismatch paradigm and a bimodal enhancement manipulation were employed. We found that LIFG is sensitive to the congruence between speech and a concurrently presented pointing gesture towards a referent, whereas the posterior STS region is not. The results of the bimodal enhancement comparisons show that the bimodal presentation of our stimuli activated similar brain areas (i.e. bilateral occipital, left premotor, and bilateral temporal incl. STS/STG/MTG) compared to the sum of the presentation of the unimodal stimuli and that these areas were not activated significantly more in the bimodal compared to the unimodal case as evident from the conjunction analysis. We will now relate these findings to the speech-gesture integration literature and previous studies of audiovisual binding more broadly.

Enhanced activation in LIFG has been found in previous studies that investigated the integration of iconic gestures with speech (e.g., Dick et al., 2014; Skipper et al., 2007; Willems et al., 2007; 2009), pantomimes with speech (Willems et al., 2009), and metaphoric gestures with speech (Kircher, Straube, Leube, Weis, Sachs, et al., 2009; Straube, Green, Bromberger, & Kircher, 2011). The common denominator in these studies is that an increase in semantic unification load led to an increase in LIFG activation (cf. Andric & Small, 2012; Dick et al., 2014; Hagoort et al., 2009; Özyürek, 2014). For instance, gestures that are unrelated to concurrently presented speech require additional semantic processing because they are harder to semantically integrate with speech compared to iconic gestures that relate to the concurrently presented speech. Therefore, the former lead to enhanced LIFG activation compared to the latter (Green et al., 2009). The same holds for metaphoric co-speech gestures compared to iconic co-speech gestures (Straube et al., 2011). Similarly, iconic gestures or pantomimes that are incongruent with speech activate LIFG more than iconic gestures and pantomimes that match the

speech they accompany (Willems et al., 2007; 2009). Confirming such previous findings, in the current study incongruence between speech and a visible object, as induced by a pointing gesture, led to enhanced activation in LIFG compared to a matched congruent condition.

Previous studies have criticized the use of mismatch paradigms in gesture-speech integration studies, for instance arguing that "mismatches, which are rarely encountered in spontaneous discourse, may trigger additional integration processes which are not normally part of multimodal language comprehension" (Holle et al., 2010, p. 876), such that activations in LIFG may be a result of "the processing of unnatural stimuli and rather relate to error detection processes" (Green et al., 2009, p. 3317; Straube et al., 2011). There are convincing reasons to believe, however, that gesture-speech mismatch manipulations tap into semantic speech-gesture integration. For instance, LIFG activation is often also present in the "match" condition compared to baseline (e.g. Willems et al., 2009). Furthermore, enhanced LIFG activation has also been found in speech-gesture integration studies that manipulated semantic load in a different way, not using a mismatch paradigm (e.g., Dick et al., 2014; Skipper et al., 2007). Dick et al. (2014), for instance, compared the integration of supplemental iconic gestures with speech to the integration of "redundant" iconic gestures with speech. The former gestures added information to the speech they accompanied (e.g. the verb in the phrase "Sparky attacked" was combined with a "peck" gesture) and therefore increased semantic processing and unification load compared to the latter gestures ("Sparky pecked" combined with a "peck" gesture). Indeed, a robust increase in activation was found in LIFG for the gestures that added information to the speech and therefore required additional semantic processing compared to the "redundant" gestures (Dick et al., 2014). Crucially, both such gestures commonly occur in everyday interactions (Holler & Beattie, 2003; Kendon, 2004; McNeill, 1992).

LIFG plays a role not only in semantic unification of speech and gesture, but also in the semantic unification of word meaning and world knowledge into a preceding context in speech itself (Hagoort, 2013; Hagoort, Hald, Bastiaansen, & Petersson, 2004; Hagoort et al., 2009; Zhu, Hagoort, Zhang, Feng, Chen, et al., 2012). The current study extends previous work in showing that semantic unification recruits LIFG across semiotic domains. LIFG thus plays a crucial role in the case of an indexical semiotic relation between gesture, speech, and a referent (the current study), in addition to symbolic and iconic manners of signification (as in arbitrary word-meaning mappings and resemblance between iconic gestures/pantomimes/pictures and referents respectively). Furthermore, a core property of language (including iconic gestures) is that is allows for displacement, i.e. the ability to refer to entities that are not immediately present (Hockett, 1960; Perniss & Vigliocco, 2014). The current study shows that also when a referent is physically present in the immediate visual context, LIFG subserves the semantic unification of auditory and visual information at a higher-order semantic level. The involvement of LIFG in the case of pointing-speech integration may be dependent on whether transmitted information is semantic and/or communicatively intended, as it was not sensitive to the congruence between speech and an attentional cue around a visual object.

Comparing the bimodal (match) condition to the sum of unimodal conditions showed that large parts of bilateral temporal cortex (including STG, STS, and MTG) and bilateral occipital cortex are also involved in speech-gesture perception in a triadic context. The bimodal enhancement manipulations arguably tap into audiovisual binding, because in the bimodal condition auditory and visual information streams are integrated whereas in the unimodal conditions they are not. Nevertheless, the regions activated by the bimodal condition showed great overlap with the regions activated by the unimodal conditions compared to baseline,

including (primary) visual and auditory cortices (see Figure 2). These findings are in line with previous studies that investigated speech-gesture integration and with studies looking at audiovisual binding more broadly using bimodal enhancement paradigms (e.g., Belardinelli et al., 2004; Calvert, Brammer, Bullmore, Campbell, Iversen, et al., 1999; Hein, Doehrmann, Müller, Kaiser, Muckli, et al., 2007; Macaluso, Frith, & Driver, 2000; Van Atteveldt, Formisano, Goebel, & Blomert, 2004; Van Atteveldt, Formisano, Blomert, & Goebel, 2007; Willems et al., 2009). Belardinelli et al. (2004), for instance, compared the perception of bimodal stimuli (pictures of entities such as animals and tools paired with environmental sounds) to the unimodal presentations of the same stimuli (audio-only, vision-only) and even found enhanced activation in visual areas in the occipital lobe in the bimodal condition. Such results indeed suggest an important role for (primary) sensory areas in audiovisual binding (cf. Van Wassenhove, Grant, & Poeppel, 2005), possibly in a dynamic interplay of several areas in a cortical network that may also involve "heteromodal" (Calvert et al., 2000) neuronal populations in STS and STG (Belardinelli et al., 2004; Calvert, 2001; Van Atteveldt et al., 2004; but see Hocking & Price, 2008).

Holle et al. (2008) argued that semantic integration of gesture and speech recruits the posterior part of the STS region. However, in both the mismatch manipulation and the bimodal enhancement comparisons, no differential activation was found in the canonical posterior portions of the STS. In contrast, Willems et al. (2009) argued that pSTS/MTG is involved in mapping different sources of information onto a common memory representation. A parsimonious interpretation of the results of the present study and these previous investigations is therefore that STS/STG may be involved in audiovisual binding/integration (cf. Beauchamp, 2005; Calvert, 2001; Dick et al., 2014), whereas MTG may subsequently map information from

different input streams onto a common representation in memory, particularly in the case of *iconic* gestures and speech (cf. Hagoort et al., 2009; Willems et al., 2009). Indeed, in audiovisual information processing the onset of enhanced activation in superior temporal areas and primary visual areas precedes the onset of enhanced activation in MTG and LIFG (Campbell, 2008; Fuhrmann Alpert, Hein, Tsai, Naumer, & Knight, 2008). Note that we do not argue that these areas are solely involved in these functions, as the same cortical area can serve different functions as a flexible component in different neural networks (e.g., Andric & Small, 2012; Hagoort, 2014; Hein & Knight, 2008; Mesulam, 1998).

In addition, we found an increase in activation in premotor areas (BA 6, left-lateralized) in the bimodal match condition compared to the sum of the unimodal conditions, and in the unimodal conditions compared to baseline. Previous studies have interpreted activity in such areas in speech-gesture comprehension in line with a putative mirror neuron system in humans (e.g., Holle et al., 2008; Skipper et al., 2007; Willems et al., 2007). Here we refrain from interpreting our results in such a way, because we did not include a condition in which participants produced pointing movements and speech themselves and because without using paradigms such as repetition suppression one cannot be sure that the activations one finds involve the same neuronal population recruited in movement planning/execution and speech production. However our findings are in line with a more general function attributed to premotor areas in the perception of purposeful hand actions (see Andric & Small, 2012), possibly related to the simulation of perceived actions, and further specify that these areas may be recruited in the perception of implied motion as well, as in the case of the static images used in the current study (cf. Pierno et al., 2009).

Finally, previous studies investigating the neural mechanisms involved in the perception of pointing gestures have focused on the gesture as a directional cue outside a speech context (see Ulloa & George, 2013). Pierno et al. (2009), for instance, compared the observation of a static image of a hand pointing at an object to the observation of a hand grasping that object and to a control condition of a hand resting next to that object. Compared to the control condition, both types of actions activated a left-lateralized network that included parietal areas (postcentral gyrus and supramarginal gyrus) and left middle occipital gyrus. The activation patterns in our visual-only condition compared to baseline (see Figure 2B) confirm that these regions are recruited in integrating pointing gestures with visual objects. In the Pierno et al. (2009) study, no area was activated significantly more in the pointing condition compared to the grasping condition. Future work may therefore investigate whether the results of the current study generalize to situations in which a speaker grasps an object while concurrently producing speech. After all, in everyday life speakers may both point at an object and grasp and hold up or place an object to bring it into their addressee's attention (Clark, 2003). It is not unlikely that the extent of overlap between pointing-speech integration and grasping-speech integration might differ as a function of the perceived communicative intentions of the speaker (see Kelly et al., 2014).

In sum, the current study investigated the neural integration of pointing gestures and speech in a visual, triadic context. Bilateral auditory and visual regions and left premotor regions were found to be involved in the concomitant perception of speech, gesture and referent, whereas LIFG subserved the semantic unification of referential gesture and speech in a triadic context. This study can be informative as a starting point for studies investigating specific populations with impairments in the comprehension and integration of deictic speech and gesture and the

subsequent establishment of joint attention in everyday life (e.g. as in autism spectrum disorders).

References

Andric, M., & Small, S. L. (2012). Gesture's neural language. *Frontiers in psychology*, *3*.

Baron-Cohen, S. (1989). Perceptual role taking and protodeclarative pointing in autism. *British Journal of Developmental Psychology*, *7*(2), 113-127.

Beauchamp, M. S. (2005). See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. *Current opinion in neurobiology*, *15*(2), 145-153.

Belardinelli, M. O., Sestieri, C., Di Matteo, R., Delogu, F., Del Gratta, C., Ferretti, A., Caulo, M., Tartaro, A., & Romani, G. L. (2004). Audio-visual crossmodal interactions in environmental perception: an fMRI investigation. *Cognitive Processing*, *5*(3), 167-174.

Boersma, P., & Weenink, D. (2009). Praat: doing phonetics by computer (Version 5.1.05) [Computer program].

Brunetti, M., Zappasodi, F., Marzetti, L., Perrucci, M. G., Cirillo, S., Romani, G. L., Pizzella, V., & Aureli, T. (2014). Do you know what I mean? Brain oscillations and the understanding of communicative intentions. *Frontiers in Human Neuroscience*, *8*, 36.

Bühler, K. (1934). *Sprachtheorie*. Jena: Fischer.

Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing.Where language, culture, and cognition meet* (pp. 9-33). Hillsdale, NJ: Erlbaum.

Call, J., & Tomasello, M. (1994). Production and comprehension of referential pointing by orangutans (Pongo pygmaeus). *Journal of Comparative Psychology*, *108*(4), 307.

Calvert, G. A. (2001). Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cerebral cortex*, *11*(12), 1110-1123.

Calvert, G. A., Brammer, M. J., Bullmore, E. T., Campbell, R., Iversen, S. D., & David, A. S.

(1999). Response amplification in sensory-specific cortices during crossmodal binding. *Neuroreport*, *10*(12), 2619-2623.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, *10*(11), 649-657.

Campbell, R. (2008). The processing of audio-visual speech: empirical and neural bases. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1493), 1001-1010.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 255, Vol. 63,* 1-174.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H. (2003). Pointing and placing. In S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet* (pp. 243-268). Hillsdale NJ: Erlbaum.

Conty, L., Dezecache, G., Hugueville, L., & Grèzes, J. (2012). Early binding of gaze, gesture, and emotion: neural time course and correlates. *The Journal of Neuroscience*, *32*(13), 4531-4539.

Cooperrider, K. (2011). Reference in action: Links between pointing and language. Doctoral dissertation, University of California, San Diego.

Dick, A. S., Mok, E. H., Beharelle, A. R., Goldin-Meadow, S., & Small, S. L. (2014). Frontal and temporal contributions to understanding the iconic co-speech gestures that accompany speech. *Human brain mapping*, *35,* 900-917.

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S.,

& Turner, R. (1995). Analysis of fMRI time-series revisited. *Neuroimage*, *2*(1), 45-53.

Friston, K. J., Holmes, A., Poline, J. B., Price, C. J., & Frith, C. D. (1996). Detecting activations in PET and fMRI: levels of inference and power. *Neuroimage*, *4*(3), 223-235.

Fuhrmann Alpert, G., Hein, G., Tsai, N., Naumer, M. J., & Knight, R. T. (2008). Temporal characteristics of audiovisual information processing. *The Journal of Neuroscience*, *28*(20), 5344-5349.

Gredebäck, G., Melinder, A., & Daum, M. (2010). The development and neural basis of pointing comprehension. *Social neuroscience*, *5*(5-6), 441-450.

Green, A., Straube, B., Weis, S., Jansen, A., Willmes, K., Konrad, K., & Kircher, T. (2009). Neural integration of iconic and unrelated coverbal gestures: a functional MRI study. *Human brain mapping*, *30*(10), 3309-3324.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in cognitive sciences*, *9*(9), 416-423.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in psychology*, *4*.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), The cognitive neurosciences, 4th ed. (pp. 819-836). Cambridge, MA: MIT Press.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438-441.

Hein, G., Doehrmann, O., Müller, N. G., Kaiser, J., Muckli, L., & Naumer, M. J. (2007). Object familiarity and semantic congruency modulate responses in cortical audiovisual integration areas. *The Journal of neuroscience*, *27*(30), 7881-7887.

Hein, G., & Knight, R. T. R. T. (2008). Superior temporal sulcus—it's my area: or is it?.

Journal of Cognitive Neuroscience, 20(12), 2125-2136.

Hockett, C. D. (1960). The origin of speech. *Scientific American, 203(3)*, 88-96.

Hocking, J., & Price, C. J. (2008). The role of the posterior superior temporal sulcus in audiovisual processing. *Cerebral Cortex*, *18*(10), 2439-2449.

Holle, H., Gunter, T. C., Rüschemeyer, S. A., Hennenlotter, A., & Iacoboni, M. (2008). Neural correlates of the processing of co-speech gestures. *Neuroimage*, *39*(4), 2010-2024.

Holle, H., Obleser, J., Rueschemeyer, S. A., & Gunter, T. C. (2010). Integration of iconic gestures and speech in left superior temporal areas boosts speech comprehension under adverse listening conditions. *Neuroimage*, *49*(1), 875-884.

Holler, J., & Beattie, G. (2003). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener?. *Gesture*, *3*(2), 127-154.

Hubbard, A. L., Wilson, S. M., Callan, D. E., & Dapretto, M. (2009). Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Human brain mapping*, *30*(3), 1028-1037.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological science*, *16*(5), 367-371.

Kelly, S., Healey, M., Özyürek, A., & Holler, J. (2014). The processing of speech, gesture, and action during language comprehension. *Psychonomic bulletin & review*, 1-7.

Kendon, A. (2004). *Gesture: Visible action as utterance.* Cambridge: Cambridge University Press.

Kircher, T., Straube, B., Leube, D., Weis, S., Sachs, O., Willmes, K., Konrad, K., & Green, A. (2009). Neural interaction of speech and gesture: differential activations of metaphoric co-verbal gestures. *Neuropsychologia*, *47*(1), 169-179.

Kita, S. (2003). *Pointing. Where language, culture, and cognition meet.* Hillsdale, NJ: Erlbaum.

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, *289*(5482), 1206-1208.

Marstaller, L., & Burianová, H. (2014). The multisensory perception of co-speech gestures–A review and meta-analysis of neuroimaging studies. *Journal of Neurolinguistics*, *30*, 69-77.

Materna, S., Dicke, P. W., & Thier, P. (2008). The posterior superior temporal sulcus is involved in social communication not specific for the eyes. *Neuropsychologia*, *46*(11), 2759-2765.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Mesulam, M. M. (1998). From sensation to cognition. *Brain*, *121*(6), 1013-1052.

Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *Neuroimage*, *25*(3), 653-660.

Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Statistical methods in medical research*, *12*(5), 419-446.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.

Özyürek, A. (2014). Hearing and seeing meaning in speech and gesture: insights from brain and behaviour. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130296.

Özyürek, A., Willems, R., Kita, S., & Hagoort, P. (2007). On-line integration of semantic

information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience, 19*(4), 605-616.

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 20130300.

Pierno, A. C., Tubaldi, F., Turella, L., Grossi, P., Barachino, L., Gallo, P., & Castiello, U. (2009). Neurofunctional modulation of brain regions by the observation of pointing and grasping actions. *Cerebral Cortex*, *19*(2), 367-374.

Poser, B. A., Versluis, M. J., Hoogduin, J. M., & Norris, D. G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, *55*(6), 1227-1235.

Quandt, L. C., Marshall, P. J., Shipley, T. F., Beilock, S. L., & Goldin-Meadow, S. (2012). Sensitivity of alpha and beta oscillations to sensorimotor characteristics of action: An EEG study of action production and gesture observation. *Neuropsychologia*, *50*(12), 2745-2751.

Sato, W., Kochiyama, T., Uono, S., & Yoshikawa, S. (2009). Commonalities in the neural mechanisms underlying automatic attentional shifts by gaze, gestures, and symbols. *NeuroImage*, *45*(3), 984-992.

Skipper, J. I., Goldin-Meadow, S., Nusbaum, H. C., & Small, S. L. (2007). Speech-associated gestures, Broca's area, and the human mirror system. *Brain and language*, *101*(3), 260-277.

Straube, B., Green, A., Bromberger, B., & Kircher, T. (2011). The differentiation of iconic and

metaphoric gestures: Common and unique integration processes. *Human brain mapping*, *32*(4), 520-533.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78,* 705-722.

Ulloa, J. L., & George, N. (2013). A Cognitive Neuroscience View on Pointing: What Is Special About Pointing with the Eyes and Hands?. *Humanamente. Journal of Philosophical Studies*, *24*, 203-228.

Van Atteveldt, N. M., Formisano, E., Blomert, L., & Goebel, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex*, *17*(4), 962-974.

Van Atteveldt, N., Formisano, E., Goebel, R., & Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron*, *43*(2), 271-282.

Van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(4), 1181-1186.

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: phonology, semantics, and sentence processing. *Neuroimage*, *30*(4), 1414-1432.

Willems, R. M., Özyürek, A., & Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cerebral Cortex*, *17*(10), 2322-2333.

Willems, R. M., Özyürek, A., & Hagoort, P. (2009). Differential roles for left inferior frontal and

superior temporal cortex in multimodal integration of action and language. *Neuroimage*, *47*(4), 1992-2004.

Zhu, Z., Hagoort, P., Zhang, J. X., Feng, G., Chen, H. C., Bastiaansen, M., & Wang, S. (2012). The anterior left inferior frontal gyrus contributes to semantic unification. *NeuroImage*, *60*(4), 2230-223

Chapter 6

Summary and Discussion

Summary and Discussion

> **"**Here a psychologist has presented his suggestions as to how in his view certain linguistic facts should be interpreted" (Bühler, 1934/1990, p. 126)

Demonstrative reference has been an important topic of investigation in many academic disciplines, including anthropology, linguistics, philosophy and psychology. However, verbal (e.g. spatial demonstratives) and non-verbal (e.g. pointing gestures) markers involved in establishing joint attention to a referent have often been studied in isolation. Furthermore, egocentric, spatial accounts of demonstrative reference have ignored that establishing joint attention to a referent is a social and collaborative process. In addition, the neural mechanisms involved in everyday multimodal reference are unclear. This thesis therefore brings spatial demonstratives and pointing gestures together using an experimental and neurobiological approach and studies the phenomenon of demonstrative reference from both the production and the comprehension side. This approach allowed for contrasting egocentric and sociocentric views of spatial deixis. In addition, it furthered our understanding of the neurobiological underpinnings that allow us to refer to the things around us and in turn those that allow us to understand other people's multimodal referential acts. Before turning to a description of the broader theoretical implications of the empirical findings presented in the current thesis, I will first briefly summarize the main results of each chapter.

Summary of the main findings

In **Chapter 2** the interplay of different social and visual contextual factors in the *production* of spatial demonstratives in Dutch and Turkish was investigated. A controlled

elicitation task was developed to disentangle the independent roles of the location of a referent, the locus of visual attention of the addressee, and the presence of a pointing gesture in demonstrative choice in two-term Dutch and three-term Turkish. In both languages, proximal demonstratives were used for referents close to the speaker and distal demonstratives for referents at three locations away from the speaker. No linear increase in distal demonstrative use was found with an increase in distance of a referent to the speaker. Moreover, the presence of visual joint attention between speaker and addressee to the referent enhanced the choice for a distal demonstrative in Turkish. Finally, pointing gestures were found to be closely tied to demonstratives but not to the use of other linguistic referential expressions. These findings were interpreted in line with a sociocentric view of demonstrative reference and against purely spatial accounts.

Chapter 3 contrasted and tested an egocentric versus a sociocentric view of demonstrative reference from a *comprehension* perspective. In two ERP experiments, participants looked at pictures in which a speaker pointed at one of two objects while they listened to her referential speech. Analysis of event-related potentials time-locked to the onset of perceiving a spatial demonstrative in such a visual context showed that, whenever participants were oriented in a dyad with the speaker in the pictures and both possible referents were between them, the comprehension of a distal demonstrative led to higher processing costs compared to the comprehension of a proximal demonstrative. This preference was irrespective of whether the referent was relatively close to the speaker or relatively far away. These findings suggest that addressees interpret spatial demonstratives as a function of the physical orientation of themselves in relation to the speaker and the spatial location of possible referents. More broadly, the findings suggest that interlocutors may construe a shared space in which all referents become

'psychologically proximal'. In general, this chapter also showed the feasibility of using neuroscientific methods to inform cognitive theory in the case of spatial demonstratives.

Chapter 4 investigated multimodal exophoric reference from the *production* side. More specifically, it examined the role of communicative intentions in shaping the kinematics of unimodal and bimodal acts of pointing as a function of the shared knowledge between speaker and addressee. Furthermore it explored how one's communicative intentions modulate the neural activity involved in planning and producing a communicative pointing gesture. Participants produced index-finger pointing gestures in the lab for a confederate addressee, while their index-finger movement kinematics and electrophysiological brain activity were continuously recorded. The informativeness of their pointing gesture and/or speech was manipulated in a block design and used as a proxy to tap into their communicative intentions. Behaviorally, participants prolonged the duration of the stroke and post-stroke hold phase of their gesture to be more informative. Electrophysiological effects of informativeness suggested that both attentional and intentional mechanisms may be involved in planning a communicative gesture. The kinematic effects are largely specific to cases where gesture carries the informational burden in a multimodal communicative act, whereas the attentional neurophysiological effect is modality-independent.

Chapter 5 investigated the neuronal mechanisms involved in the integration of pointing gestures and referential speech in *comprehension* using fMRI. Participants watched images of a person pointing at one of two objects and listened to her speech. As evident from a mismatch-match comparison, left inferior frontal gyrus (LIFG) is recruited in semantic unification of referential speech and pointing gesture towards an object. A bimodal enhancement manipulation showed the involvement of bilateral temporal areas (STG/STS/MTG), occipital areas, and

premotor areas in the perception and integration of speech and gesture in a referential, triadic context. These results specify the nodes in the network involved in the perception and understanding of referential speech-gesture combinations. At the same time, they suggest overlap in the mechanisms involved in integration and unification of different types of gesture (e.g., iconic and pointing gestures) with speech. These findings may be taken as a starting point in the study of populations that have impairments related to multimodal communication, such as in autism spectrum disorder.

No more egocentricity! Towards a sociocentric view of spatial deixis

One aim of the current thesis was to contrast and test egocentric and sociocentric views of spatial deixis. The findings presented in Chapters 2 and 3 falsify three claims that are central to the egocentric, spatialist account of spatial deixis, which has been the dominant view in the literature over the last 80 years:

1. "[T]he anchoring point of deictic expressions is egocentric (or, better, speaker-centric). Adult speakers skillfully relate what they are talking about to this me-here-now" (Levelt, 1989, p. 46).
2. By using spatial demonstratives, speakers "indicate the relative distance of an object, location, or person vis-à-vis the deictic center (…), which is usually associated with the location of the speaker" (Diessel, 1999, p. 36).
3. "[D]emonstratives are interpreted based on the speaker's body" (Diessel, 2014, p. 122).

The first claim is seriously challenged by the data presented in Chapter 2. In Turkish, speakers took into account the locus of visual attention of the addressee when using a demonstrative. The distal demonstrative term *o* was used significantly more often when there was joint attention between speaker and addressee to a referent compared to when the addressee looked at a different part of the visual scene. This finding is incompatible with an egocentric view of deixis in which the speaker is the sole anchoring point and in which factors related to the addressee do not play a role in driving a speaker's demonstrative choice. Rather than relating the demonstrative they produced to the me-here-now, participants related it to the *we*-here-now, taking into account the addressee.

The second claim, which identifies the relative distance of a referent to the speaker as the main factor driving demonstrative choice, is also not in line with the data reported in this thesis. In Chapter 2, the relative distance of a referent to the speaker did not drive demonstrative choice in Dutch or Turkish. Although in both languages proximal demonstratives were used for referents close to the speaker and distal terms for referents further away, there was no linear increase of distal demonstrative use with a linear increase in distance of a referent to the speaker. Thus, *relative distance* does not drive demonstrative choice. Rather, speakers may carve up the physical space into different meaningful zones (Enfield, 2003; Kendon, 1977). This suggests that the *location* of a referent in a specific region of meaningful space is more important than its relative distance to the speaker. This finding confirms a similar conclusion drawn by Enfield (2003) on the basis of analysis of observational data from Lao. Moreover, contextual factors beyond the location of a referent also influence the choice for one demonstrative over another, such as the locus of attention of the addressee. Because the pointing gesture in exophoric use will generally create a vector towards the referent, and the height of the raised arm indicates the

distance of the referent (e.g., Gonseth, Vilain, & Vilain, 2013), one could even argue that it is the pointing gesture, rather than the demonstrative term, that indicates the relative distance of a referent to speaker and addressee.

The third claim is directly falsified by the findings presented in Chapter 3. The participant addressees in two experiments did not interpret demonstratives on the basis of the speaker's body, but rather as a function of the orientation of their own body in relation to the speaker's body and the location of referents inside or outside of the dyad. These findings again confirm that demonstrative reference is a sociocentric phenomenon in which both speaker and addressee play a crucial role. Interestingly, these findings were not predicted from linguistic intuitions elicited in a pre-test. The discrepancy between intuitions and actual data is reminiscent of similar findings in the domain of demonstrative production (e.g. Enfield, 2003; Özyürek, 1998). People's personal metalinguistic intuitions about their demonstrative choice and their demonstrative comprehension align with the egocentric, spatialist view, but not with patterns of use and perception in actual data.

The findings in Chapters 2 and 3 together suggest that people may carve up space in different ways in different contexts (cf. Enfield, 2003). In Chapter 2, a zone close to the speaker was distinguished from the rest of the physical space. In Chapter 3, the shared space between speaker and addressee was distinguished from the space outside of the dyad. Future work may shed more light on which physical and social factors influence how interlocutors carve up the space around them in the context of demonstrative reference. In addition to physical boundaries and social considerations, the interlocutors' areas of manual and attentional engagement during interaction, and whether these overlap or not, may play a crucial role in the process of building

up shared space (Enfield, 2003; Hanks, 1990), subsequently influencing their choice of demonstrative.

Supporting evidence against egocentric and purely spatial accounts of demonstrative reference comes from research looking at pointing gestures and from research into referential communication more broadly.

Supporting evidence from the gestural modality

In everyday communication the production of a spatial demonstrative is generally part of a richer communicative act that involves a pointing gesture produced by the speaker. The kinematic findings presented in Chapter 4 indicate that people tailor the form properties of their gesture to the context-specific needs of their addressee, which further specifies previous observational work (Enfield, Kita, & De Ruiter, 2007; Kendon & Versante, 2003). This suggests that people's sociocentric approach to demonstrative reference expresses itself in both the spoken and the gestural modality. The gestural findings align well with the broader literature on pointing, for instance related to the gesture's role in ontogeny.

An important focus over the last few decades in the study of pointing gestures has been on the gestures' role in prelinguistic communication and first language acquisition (e.g., Bates, Camaioni, & Volterra, 1975; Butcher & Goldin-Meadow, 2000; Butterworth, 2003; Csibra, 2010; Iverson & Goldin-Meadow, 2005; Kita, 2003; Leung & Rheingold, 1981; Moore & D'Entremont, 2001; Tomasello, 2008; Tomasello, Carpenter, & Liszkowski, 2007). By pointing to the things around them infants start expressing their *communicative intentions* and they continue to do so throughout life. The aim of infants' (and adults') pointing gestures is often simply declarative, i.e. to share interest in a certain referent and for the addressee to recognize

one's communicative intentions (Tomasello et al., 2007). Even in imperative pointing, generally defined as pointing to request an object from someone, it seems reasonable that one has to recognize the other as an intentional, social agent (Southgate, van Maanen, & Csibra, 2007). It is hard to unite such a view of pointing as deeply social and communicative with an egocentric view of spatial deixis in which the speaker disregards the addressee in choosing a demonstrative and only takes into account the relative distance of an object to him- or herself. Rather, the social and communicative nature of human pointing confirms that multimodal demonstrative reference is an interpersonal, collaborative process in which the addressee plays a pivotal role (Clark & Bangerter, 2004).

What is the nature of the exact interplay between spatial demonstratives and pointing gestures in sociocentric demonstrative reference? Diessel (2006) suggests that demonstratives and pointing gestures have the same function in shifting the addressee's attention to an intended referent. However, simply uttering a demonstrative term without any indexical bodily accompaniment will generally not suffice in making one's addressee shift gaze and identify an intended referent, whereas a pointing gesture without concurrently produced speech may suffice in certain cases. Alternatively, it has been suggested that demonstratives shift the addressee's attention to the speaker's pointing gesture (Bangerter, 2004; Bühler, 1934), while "pointing gestures immediately deflect that attention elsewhere" (Cooperrider, 2011, p. 7), i.e. to the referent. As outlined in Chapter 4, people design the kinematics of their gesture for their addressee, and the use of a demonstrative could indeed make the addressee pay attention to such an effort. In addition, the specific demonstrative term used will provide some additional information to the addressee about where the intended referent is located (such as inside or

outside of the shared space). In such a sociocentric view, the speaker thus uses both the gestural and the spoken modality to inform the addressee.

Supporting evidence from a broader perspective

The current thesis rejects the egocentric account of spatial deixis on empirical grounds. Moreover, this account also has undesirable philosophical shortcomings. For instance, deictic terms almost always have a relational character. The words *here, now,* and *I*, for example, derive their meaning for a large part because of their opposition to other terms - *there, then,* and *you* (Jones, 1995). In order then to assume a deictic center, one has to presuppose "the existence and identification of entities – people, places and times – outside the charmed circle of individual subjectivity" (Jones, 1995, p. 33). Consequently, the speaking ego always stands and derives its meaning in relation to a hearing other, and deictic terms, including *this* and *that*, always already assume a social, relational context in which the speaker is just one element of a larger deictic field. It is exactly in such a social situation that spatial deictic reference usually takes place, and the addressee is just as necessary and important in such a field as the speaker.

The theoretical progress from an egocentric to a sociocentric view of spatial deixis is reminiscent of the development in our understanding of reference production more broadly. In the General Introduction I have argued that the egocentric account is largely based on linguistic intuitions and not on rigorous empirical testing. Sophisticated observational work in different languages suggested that these intuitions were unreliable, which was confirmed experimentally in the current thesis. As outlined by Clark and Bangerter (2004), similar developments have taken place in the field of reference production in general. Traditional accounts of reference production considered reference production an autonomous and addressee-blind act that speakers do on their own without taking into account beliefs about their addressees (Clark & Bangerter,

2004, pp. 26-27). However, more recent views consider reference production a collaborative act that requires that speaker and addressee work together, for instance in building conceptual pacts (e.g., Clark & Wilkes-Gibbs, 1986), i.e. bilateral agreements on how to conceptualize and name a particular referent (Brennan & Clark, 1996). Such agreement is established through interaction, and again, the addressee is just as important as the speaker in reaching agreement and establishing reference. A sociocentric view of spatial deictic reference thus fits well within the broader context of reference production as a social, interactive phenomenon.

Towards a neurobiological account of demonstrative reference

The results presented in Chapters 3-5 further our understanding of the neural and cognitive mechanisms involved in the production and comprehension of demonstrative reference. In the planning of referential speech and gesture, attentional resources are recruited as a function of one's communicative intentions. Furthermore, the production of declarative pointing involves activation in the mentalizing network in adults (cf. Brunetti et al., 2014) and previous work has shown that it activates its precursor in infants (Henderson et al., 2002). In comprehension, hearing an (at the pragmatic level) incorrectly used spatial demonstrative term leads to an ERP component that is similar to the "typical" N400 in its negative directionality and its time-course. Moreover, the integration of referential speech and gesture in comprehension requires audiovisual and semantic integration processes to take place at a neuronal level.

Together these findings thus suggest a complex interplay between different neural mechanisms involved in the production and comprehension of multimodal (i.e. pointing gestures and speech) reference, including attentional (as indicated by the P3b effect in Chapter 4), intentional (the frontal ERP effect in Chapter 4), action-related (the readiness potential in Chapter 4 and the activation of premotor areas in Chapter 5), audiovisual perceptual and

integrative (primary visual and auditory, and temporal lobe activations in Chapter 5), and word processing and semantic unification mechanisms (N400-type effects in Chapter 3 and LIFG activation in Chapter 5). A challenge for future work will be to further relate the activation of different networks and different nodes in these networks to their functional roles over time and to identify possible overlap between neuronal populations involved in reference production and comprehension.

Future directions

The work reported in this thesis focused on the *exophoric* gestural use of spatial demonstratives, i.e. when a speaker uses a demonstrative in reference to a certain entity (such as a physical object) that is physically present in the extra-linguistic, physical context of a linguistic interaction (Halliday & Hasan, 1976; see Levinson, 2004). However, demonstratives serve several functions in language, for instance non-deictically in anaphoric and empathetic uses (see Diessel, 1999; Himmelmann, 1996; Levinson, 2004). Furthermore, the empirical study of demonstratives as described in this thesis was restricted to their adnominal/adjectival use (as in *this book*). Whether the conclusions drawn in the current thesis generalize to the use of demonstratives as pronouns (as in *this tastes delicious*) or adverbs (e.g., *here-there*) is left for future study.

Similarly, the current thesis also only investigated index-finger pointing gestures in their exophoric use. In addition, it focused on simple situations in which *what was pointed at* was also *what was referred to*. In everyday life, however, pointing gestures may be more or less abstract. By using abstract pointing gestures, speakers place a concept or idea in physical space, and as such they do not direct the addressee's attention to a physically present object (McNeill, 1992). Moreover, pointing gestures often require quite some inferencing in order to be fully understood

(Clark, Schreuder, & Buttrick, 1983). For instance, depending on the common ground between interlocutors a simple pointing gesture produced towards a particular building may mean something complex like "Remember the last time we were here ten years ago" (see Clark et al., 1983; Tomasello, 2008, for better examples). Future work may shed more light on whether the findings presented in this thesis generalize to more complex uses of pointing gesture, such as when there is a discrepancy between *demonstratum* and referent (Clark, 1996; Clark et al., 1983; Clark & Bangerter, 2004).

Conclusion

This thesis focused on the production and comprehension of spatial demonstrative terms and index-finger pointing gestures as pivotal parts of the larger complex system of referential communication in a visual, triadic context. In the General Introduction, I argued that seemingly simple acts of reference require a complex interplay between speaker and addressee, arguably relying on multiple cognitive mechanisms and multimodal social cues: In a prototypical instance of successful everyday referential communication, a speaker produces a manual pointing gesture to a physical object, often in temporal alignment with a spoken referential expression that canonically contains a spatial demonstrative (as in *I have bought <u>that</u> book*), while alternating gaze between addressee and referent. At the same time, the addressee perceives the speech, gesture, and other bodily behavior of the speaker, integrates the transmitted visual and auditory information, recognizes and understands the speaker's communicative intention and social motive, and shifts her gaze to identify the referent and establish joint attention.

The four empirical chapters of this thesis each focused on different elements of this complex phenomenon. It was found that speakers may take into account the visual attentional status of their addressee in their choice of demonstrative (Chapter 2) and tailor the kinematics of

their pointing gestures to the needs of their addressee in line with their communicative intentions, and supported by attentional and intentional underlying neural mechanisms (Chapter 4). Furthermore, speaker and addressee may build up a shared space in which all possible referents become "psychologically proximal" (Chapter 3). Understanding a speaker's referential speech and gesture recruits a neural network that comprises left inferior frontal, bilateral temporal and occipital, and left premotor areas (Chapter 5).

All in all, in line with supporting evidence from studies of multimodal reference production in general, the findings presented in this thesis suggest that it is now time to definitively leave behind egocentric and purely spatial accounts of demonstrative reference, which have dominated the field for at least the last 80 years. Furthermore, the results of this thesis open up new avenues towards understanding the neural and cognitive mechanisms underlying demonstrative reference from a social and multimodal perspective.

References

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*(6), 415-419.

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development*, 205-226.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482.

Brunetti, M., Zappasodi, F., Marzetti, L., Perrucci, M. G., Cirillo, S., Romani, G. L., Pizzella, V., & Aureli, T. (2014). Do you know what I mean? Brain oscillations and the understanding of communicative intentions. *Frontiers in Human Neuroscience*, *8*, 36.

Bühler, K. (1934). *Sprachtheorie*. Jena: Fischer.

Butcher, C. & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In D. McNeill (Ed.), *Language and gesture* (pp. 235-257). New York: Cambridge University Press.

Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing.Where language, culture, and cognition meet* (pp. 9-33). Hillsdale, NJ: Erlbaum.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. H., & Bangerter, A. (2004). Changing ideas about reference. In I. A. Noveck, & D. Sperber (Eds.), *Experimental Pragmatics* (pp. 25-49). Basingstoke: Palgrave Macmillan.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *Journal of verbal learning and verbal behavior*, *22*(2), 245-258.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1-39.

Cooperrider, K. (2011). Reference in action: Links between pointing and language. Doctoral dissertation, University of California, San Diego.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind & Language*, *25*(2), 141-168.

Diessel, H. (1999). *Demonstratives. Form, Function, and Grammaticalization*. Amsterdam: John Benjamins.

Diessel, H. (2006). Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics*, *17*(4), 463–489.

Diessel, H. (2014). Demonstratives, Frames of Reference, and Semantic Universals of Space. *Language and Linguistics Compass*, *8*(3), 116-132.

Enfield, N. J. (2003). Demonstratives in space and interaction: Data from Lao speakers and implications for semantic analysis. *Language*, 82-117.

Enfield, N. J., Kita, S., & De Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, *39*(10), 1722-1741.

Gonseth, C., Vilain, A., & Vilain, C. (2013). An experimental study of speech/gesture interactions and distance encoding. *Speech communication*, *55*(4), 553-571.

Halliday, M. A. K., & Hasan, R. (1977). *Cohesion in English*. London, UK: Longman Group Ltd.

Hanks, W. F. (1990). *Referential practice: Language and lived space among the Maya*. Chicago: University of Chicago Press.

Henderson, L. M., Yoder, P. J., Yale, M. E., & McDuffie, A. (2002). Getting the point: Electrophysiological correlates of protodeclarative pointing. *International Journal of Developmental Neuroscience*, *20*(3), 449-458.

Himmelmann, N. (1996). Demonstratives in narrative discourse: A taxonomy of universal uses. In B. Fox (Ed.), *Studies in anaphora* (pp. 205-254). Amsterdam: John Benjamins.

Iverson, J. M., & Goldin-Meadow, S. (2005). Gesture paves the way for language development. *Psychological science*, *16*(5), 367-371.

Jones, P. (1995). Philosophical and theoretical issues in the study of deixis: A critique of the standard account. In K. Green (Ed.), *New essays in Deixis: Discourse, narrative, literature* (pp. 27-48). Amsterdam: Rodopi.

Kendon, A. (1977). Spatial organization in social encounters: The F-formation system. In A. Kendon (Ed.), *Studies in the behavior of social interaction* (pp. 179-208). Lisse: Peter de Ridder Press.

Kendon, A. & Versante, L (2003). Pointing by hand in "Neapolitan". In S. Kita (Ed.), *Pointing. Where language, culture, and cognition meet* (pp. 109-137). Hillsdale NJ: Erlbaum.

Kita, S. (2003). *Pointing. Where language, culture, and cognition meet.* Hillsdale, NJ: Erlbaum.

Leung, E. H., & Rheingold, H. L. (1981). Development of pointing as a social gesture. *Developmental Psychology*, *17*(2), 215.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: Bradford.

Levinson, S. C. (2004). Deixis. In L. Horn (Ed.), *The handbook of pragmatics* (pp. 97-121). Oxford: Blackwell.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Moore, C., & D'Entremont, B. (2001). Developmental changes in pointing as a function of attentional focus. *Journal of Cognition and Development*, *2*(2), 109-129.

Özyürek, A. (1998). An analysis of the basic meaning of Turkish demonstratives in face-to-face

conversational interaction. In S. Santi, I. Guaitella, C. Cave, & G. Konopczynski (Eds.), *Oralité et gestualité: Communication multimodale, interaction: actes du colloque ORAGE 98* (pp. 609-614). Paris: L'Harmattan.

Southgate, V., Van Maanen, C., & Csibra, G. (2007). Infant pointing: Communication to cooperate or communication to learn?. *Child development*, *78*(3), 735-740.

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78,* 705-722.

Nederlandse samenvatting

In het dagelijks leven verwijzen we naar de dingen in de wereld om ons heen. We gebruiken het gesproken woord om dingen te benoemen, en onze lichamen (in het bijzonder ogen, hoofd, handen en torso) om een locatie of entiteit aan te duiden waarop een gesprekspartner zijn aandacht dient te vestigen om te herkennen waar we het precies over hebben (Bühler, 1934; Clark & Bangerter, 2004). Op het eerste gezicht lijkt dit een erg simpel onderdeel van menselijke communicatie. Nadere inspectie leert, echter, dat simpele referentiële taaluitingen een complexe interactie vereisen tussen de spreker en zijn/haar gesprekspartner, waarbij meerdere cognitieve mechanismen en multimodale, sociale overwegingen een rol spelen. In een prototypisch voorbeeld van succesvolle referentiële communicatie maakt een spreker een wijsgebaar richting een fysiek object, vaak in temporele afstemming met een gesproken referentiële uitdrukking (zoals *Ik heb zojuist <u>dat</u> boek gekocht*) inclusief aanwijzend voornaamwoord, ondertussen de blik alternerend tussen gesprekspartner en het object waarnaar verwezen wordt (de 'referent'). Tegelijkertijd hoort de gesprekspartner de spraak, ziet zij de gebaren en andere lichamelijke uitingen van de spreker, integreert zij de overgebrachte visuele en auditieve informatie, herkent zij de communicatieve intentie en sociale motieven van de spreker, en verplaatst zij haar blik richting de referent om daar, samen met de spreker, haar aandacht even op te rusten. Dit proefschrift heeft als doel een beter begrip te krijgen van zulke complexe situaties waarin een spreker voor een adressant met taal en gebaar verwijst naar een zichtbaar object. Het neemt hierin een experimentele, cross-linguïstische en neurobiologische benadering. Iedere analytische benadering van een complex fenomeen vraagt om het opdelen van het complexe geheel in simpelere, behapbare elementen (Kita, 2003). De verschillende hoofdstukken van het proefschrift richten zich daarom op verschillende onderdelen van het

hierboven beschreven fenomeen van referentiële communicatie in taal en gebaar, om de elementen vervolgens in een algemene discussie terug bijeen te brengen.

Op basis van eerder onderzoek binnen de linguïstiek (Enfield, 2003; Levinson, 1983; Lyons, 1977), de antropologie (Hanks, 1990), de filosofie (Peirce, 1955; Russell, 1940; Quine, 1960; Wittgenstein, 1953) en de psychologie (Clark & Sengul, 1978; Coventry et al., 2014; Tomasello et al., 2007) kan er grofweg een onderscheid gemaakt worden tussen twee theoretische benaderingen in het denken over verwijzen in taal en gebaar. Aanhangers van de eerste stroming zijn vaak geïnspireerd door het werk van Karl Bühler (1934) en stellen, impliciet dan wel expliciet, dat verwijzen een egocentrische aangelegenheid is. Deze stroming richt haar aandacht voornamelijk op het ontdekken en beschrijven van patronen in het gebruik van aanwijzende voornaamwoorden, zoals *dit* en *dat* in het Nederlands. Door zulke woorden te gebruiken zouden sprekers de relatieve afstand van een bepaald object, een locatie, of een persoon aangeven in verhouding tot hun eigen fysieke locatie. In de praktijk van het Nederlands zou dat betekenen dat sprekers *dit boek* zeggen wanneer ze verwijzen naar een boek dat relatief dicht bij hen in de buurt verkeert, en *dat boek* voor een boek dat zich relatief ver weg bevindt. De spreker, met andere woorden, schikt zich in de rol van ego en relateert alles aan zijn of haar eigen gezichtspunt (Lyons, 1977). Er wordt zelfs beweerd dat sprekers van *alle* talen een dergelijk egocentrische benadering gebruiken wanneer ze met aanwijzende voornaamwoorden verwijzen naar de dingen om zich heen (Diessel, 2004). Deze egocentrische benadering van het verwijzen is intuïtief erg aannemelijk en erg invloedrijk in de wetenschappelijke literatuur over het onderwerp (zie bijvoorbeeld Anderson & Keenan, 1985; Clark & Sengul, 1978; Coventry, Valdés, Castillo, & Guijarro-Fuentes, 2008; Diessel, 2005, 2014; Fillmore, 1982; Halliday & Hasan, 1977; Hottenroth, 1982; Lakoff, 1974; Levelt, 1989; Lyons, 1977; Rauh, 1983; Russell,

1940; Stevens & Zhang, 2013). Een dergelijke, egocentrische benadering voorspelt dat sprekers niet alleen hun talige uitingen, maar ook hun wijsbewegingen niet aanpassen aan de contextuele behoeften van hun gesprekspartner.

Een alternatief op de egocentrische benadering is, per definitie, de stelling dat verwijzen een sociocentrische aangelegenheid is (Hanks, 1992). In een dergelijke benadering staat niet enkel de spreker centraal, maar is degene voor wie de spreker verwijst (de adressant) minstens even belangrijk. Toegepast op het gebruik van aanwijzende voornaamwoorden betekent dit dat de keuze voor een specifiek aanwijzend voornaamwoord niet gedreven wordt door de relatieve afstand van het object waarnaar verwezen wordt in verhouding tot de spreker, maar veeleer door contextuele factoren die te maken hebben met de verhouding tussen spreker, adressant en object. Enfield (2003) beweert bijvoorbeeld op basis van uitgebreide analyse van referentieel taalgebruik in alledaagse contexten dat verwijzen een sociaal, interactief proces is waarin de keuze voor een bepaald aanwijzend voornaamwoord gedreven wordt door hoe gesprekspartners de fysieke ruimte om zich heen waarnemen en interpreteren. Jungbluth (2003) stelt eveneens dat het gebruik van aanwijzend voornaamwoorden toont dat verwijzen geen egocentrische aangelegenheid is. Haar bevindingen suggereren dat gesprekspartners, wanneer ze in een gesprek tegenover elkaar zitten, een onderscheid maken tussen de ruimte tussen hen in en de ruimte buiten de conversationele dyade. Alle ruimte binnen de dyade wordt als 'dichtbij' ervaren, en de keuze voor een bepaald aanwijzend voornaamwoord zou in dergelijke gevallen onafhankelijk zijn van het gegeven of een object zich relatief dichtbij of op een grotere relatieve afstand van de spreker binnen de dyade bevindt. Een dergelijke, sociocentrische benadering voorspelt dat sprekers niet alleen hun talige uitingen, maar ook hun wijsbewegingen kinematisch afstemmen op de communicatieve behoeften van hun gesprekspartner.

Dit proefschrift test en contrasteert in vier experimentele hoofdstukken de egocentrische en de sociocentrische benaderingen van verwijzen in taal en gebaar. Het kiest voor een experimentele benadering en is daarmee een aanvulling op linguïstisch en antropologisch veldwerk dat elders gedaan is (bijvoorbeeld Enfield, 2003; Hanks, 1990). Desalniettemin wordt het gebruik en begrip van taal en gebaar in referentiële communicatie in dit proefschrift in een rijke, multimodale context bestudeerd. Eerder onderzoek naar aanwijzend voornaamwoorden heeft zich over het algemeen voornamelijk bezig gehouden met de talige elementen van referentiële communicatie, daarbij minder aandacht schenkend aan de handgebaren die vaak gepaard gaan met een succesvolle verwijs-act. Eerder onderzoek naar wijsgebaren belichtte juist vaak enkel de non-verbale kant van het fenomeen. Dit proefschrift brengt taal en gebaar bij elkaar. De experimentele methoden gebruikt in het beschreven onderzoek, zoals *elektro-encefalografie* (EEG) en *functional magnetic resonance imaging* (fMRI), worden ontleend aan de cognitieve neurowetenschappen. Deze benadering heeft twee voordelen. Ten eerste verschaft het de mogelijkheid om de twee theoretische benaderingen te testen door naar de respons van het brein te kijken op verschillende talige en non-verbale uitingen. Ten tweede zijn de cognitieve en neurale mechanismen betrokken bij het produceren en begrijpen van referentiële communicatieve uitingen verre van bekend en brengt dit proefschrift ons denken daarover een stap verder.

Hoofdstuk 2 van het proefschrift rapporteert een onderzoek naar de invloed van contextuele factoren op het gebruik van aanwijzend voornaamwoorden in het Nederlands en het Turks. Zowel Nederlandse als Turkse proefpersonen kregen visuele stimuli te zien waarin een spreker, een adressant en een object aanwezig waren. Er werd hun gevraagd inleidende zinnen af te maken op basis van de gepresenteerde visuele context. In de context werden de locatie van het

object t.o.v. de spreker en de adressant, de visuele focus van de aandacht van de adressant, en het gebruik van een wijsgebaar door de spreker orthogonaal gemanipuleerd. Het gebruik van aanwijzend voornaamwoorden in het Nederlands (2-term-systeem) en het Turks (3-term-systeem) werd geanalyseerd in verhouding tot het gebruik van lidwoorden in het Nederlands en tot het gebruik van een bredere klasse van bepaalde uitdrukkingen in het Turks. In beide talen werden *proximale* aanwijzend voornaamwoorden (*dit/deze* in het Nederlands, *bu* in het Turks) gebruikt voor objecten dicht bij de spreker. *Distale* aanwijzend voornaamwoorden (*dat/die* in het Nederlands, *şu, o* in het Turks) werden in beide talen gebruikt voor objecten op drie locaties iets verder verwijderd van de spreker. Er werd geen lineaire toename in het gebruik van distale aanwijzend voornaamwoorden gevonden met een lineaire toename van de fysieke afstand van een object tot de spreker. Dit gaat in tegen de predicties van de egocentrische stroming, die een dergelijk lineair effect voorspelt. Het Turkse aanwijzend voornaamwoord *o* werd relatief vaker gebruikt wanneer de aandacht van spreker en adressant reeds op het referent-object rustte dan wanneer dit niet het geval was. Wijsgebaren waren in beide talen sterk verbonden met het gebruik van aanwijzend voornaamwoorden, maar niet met het gebruik van andere referentiële uitdrukkingen zonder aanwijzend voornaamwoord. Deze bevindingen tonen aan dat de keuze van een specifiek aanwijzend voornaamwoord niet enkel gedreven wordt door de relatieve afstand van een object tot de spreker. Ook overwegingen die te maken hebben met de aandachtsfocus van de adressant en het parallelle gebruik van een wijsgebaar spelen een rol.

Hoofdstuk 3 van het proefschrift contrasteerde de egocentrische met de sociocentrische benadering tot verwijzen in taal en gebaar in twee EEG experimenten in het Nederlands. Proefpersonen kregen plaatjes te zien met daarop een spreker die een wijsgebaar maakte naar een object terwijl ze tegelijkertijd luisterden naar zinnen uitgesproken door de spreker. Analyse van

de elektrofysiologische respons op het horen van een aanwijzend voornaamwoord in een dergelijke audiovisuele context toonde aan dat het verwerken van distale aanwijzend voornaamwoorden (*die/dat*) hogere kosten met zich meebracht dan het verwerken van proximale aanwijzend voornaamwoorden (*deze/dit*) wanneer een object binnen de conversationele dyade tussen spreker en proefpersoon gesitueerd was. Deze verwerkingskosten waren onafhankelijk van de locatie van het object binnen de dyade, relatief dichtbij of verder weg van de spreker. Deze bevindingen tonen aan dat adressanten aanwijzend voornaamwoorden interpreteren op basis van de fysieke oriëntatie van henzelf in verhouding tot de spreker en de locatie van mogelijke referenten. In een breder kader suggereert dit dat gesprekspartners een gedeelde ruimte opbouwen tijdens een conversatie, waarbij alle objecten binnen die ruimte als psychologisch dichtbij worden ervaren. Eveneens toont dit onderzoek aan dat het haalbaar is om methoden uit de cognitieve neurowetenschap te gebruiken om cognitieve theorieën naar verwijzen in taal en gebaar verder te brengen.

Hoofdstuk 4 van dit proefschrift onderzoekt hoe iemands communicatieve intenties de kinematische parameters van zijn/haar wijsgebaren beïnvloeden. Eveneens bestudeert het hoe iemands communicatieve intenties de neurale activiteit beïnvloeden die voorafgaat aan het produceren van een communicatief wijsgebaar. Proefpersonen werd gevraagd om in het lab wijsgebaren te maken voor een adressant terwijl een sensor op de nagel van hun wijsvinger de kinematische kenmerken van hun wijsgebaar registreerde en tegelijkertijd hun elektro-encefalogram continue werd geregistreerd. In een block design werd de informatieve waarde van de wijsbewegingen gemanipuleerd, waardoor proefpersonen in verschillende condities een verschillende communicatieve intentie hadden. Proefpersonen vertraagden hun wijsbeweging wanneer ze de intentie hadden om informatiever te zijn, en ze verlengden de duur van de fase

waarin de vinger gericht op een referent stil werd gehouden om de adressant meer tijd te gunnen om de juiste referent te herkennen. De elektrofysiologische bevindingen suggereerden dat proefpersonen meer aandacht besteden aan de taak wanneer ze informatievere wijsbewegingen maakten, en dat frontale, intentionele mechanismen actiever waren bij het plannen van een informatiever wijsgebaar.

Hoofdstuk 5 van dit proefschrift onderzoekt met behulp van fMRI de neuronale mechanismen die betrokken zijn bij het integreren van spraak en wijsgebaar in een alledaagse, referentiële, audiovisuele context waarin een spreker in woord en gebaar verwijst naar een object. Proefpersonen keken in de MRI-scanner naar plaatjes van een spreker die naar een object wees terwijl ze luisterden naar haar spraak. Analyse van de op deze manier verkregen data toonde aan dat het inferieure deel van de frontale cortex in de linker hersenhelft betrokken is bij de semantische integratie van woord en gebaar. Voorts werd duidelijk dat, meer algemeen, temporale, occipitale en premotor gebieden betrokken zijn bij het waarnemen en integreren van de auditieve en visuele informatie die wordt overgebracht wanneer iemand naar een object verwijst in taal en gebaar. De bevindingen die het belang aantonen van het inferieure deel van de frontale cortex tonen samen met eerder onderzoek naar de integratie van andere typen gebaren (bijvoorbeeld *iconische* gebaren) met spraak aan dat dit gedeelte van ons brein van belang is bij de semantische integratie van verschillende informatiestromen.

Op basis van de vier experimentele hoofdstukken van dit proefschrift kunnen meerdere conclusies getrokken worden wat betreft de theoretische stromingen met betrekking tot verwijzen in taal en gebaar en wat betreft de neurobiologische structuren en mechanismen die het ons mogelijk maken om naar de dingen in de wereld om ons heen te verwijzen en zulke verwijzingen te begrijpen. Hoofdstukken 2 en 3 tonen aan dat het verwijzen naar een referent

met behulp van een aanwijzend voornaamwoord niet simpelweg door egocentrische overwegingen wordt gedreven, zoals decennialang in de linguïstische, filosofische en psychologische literatuur is beweerd. Hoofdstuk 3 toont zelfs aan dat de relatieve afstand van een object tot de spreker geen enkele rol lijkt te spelen in de interpretatie van een aanwijzend voornaamwoord. Dergelijke bevindingen suggereren dat het tijd is om de egocentrische stroming te vervangen door een theorie waarin centraal staat dat zowel spreker als adressant een cruciale rol spelen in situaties waarin iemand naar iets in de wereld verwijst. Zo'n theorie wordt bevestigd door de bevindingen uit Hoofdstuk 4, waarin duidelijk werd dat proefpersonen de specifieke kinematische kenmerken van hun wijsbeweging afstellen op de informatieve behoeften van hun adressant. Wederom toont dit dus aan dat het verwijzen naar een object geen egocentrisch proces is, maar veeleer een sociale aangelegenheid waarbij spreker en adressant samen betrokken zijn en samenwerken om tot een staat van gedeelde aandacht te komen. De gemiddelde dreumes verwijst al rond zijn/haar eerste verjaardag met een wijsgebaar naar objecten in zijn/haar omgeving, vaak simpelweg *declaratief* om interesse in een dergelijk object te delen en om zijn/haar communicatieve intenties kenbaar te maken aan een ander. Het is lastig om een dergelijk proces dat al van jongs af aan sociaal-communicatief en collaboratief is te beschouwen in het daglicht van een egocentrische theorie van verwijzen in taal en gebaar.

Het onderzoek beschreven in dit proefschrift zet ook een stap in de goede richting wat betreft het begrijpen van de neurobiologische en cognitieve mechanismen die betrokken zijn bij het produceren en begrijpen van verwijzingen in taal en gebaar. Tijdens het plannen van een wijsgebaar worden attentionele bronnen aangesproken op basis van iemands communicatieve intenties. Het horen van een aanwijzend voornaamwoord dat op een pragmatisch niveau niet overeenkomt met de verwachtingen die men heeft op basis van de positie van menzelf ten

opzichte van de spreker leidt tot een ERP component die in timing en directionaliteit vergelijkbaar is met de "reguliere" N400 component. De audiovisuele perceptie en semantische integratie van taal en wijsgebaar teneinde een multimodale verwijzing te begrijpen leidt tot neurale activiteit in frontale, temporele, occipitale en premotor gebieden van ons brein.

Kortom, het onderzoek beschreven in dit proefschrift toont aan dat het tijd is om de egocentrische theorie van verwijzen in taal en gebaar te vervangen door een sociaal alternatief. Verder opent het nieuwe deuren richting een beter begrip van de cognitieve en neurale mechanismen die ten grondslag liggen aan het produceren en begrijpen van verwijzingen in taal en gebaar - een alledaags, sociaal fenomeen dat fundamenteel is in het tot stand brengen van geslaagde menselijke communicatie.

Acknowledgments

The work reported in this thesis has benefitted a lot from the help of many people. First and foremost, I would like to thank my doctoral supervisors, Professors Aslı Özyürek and Peter Hagoort. It is a great privilege to work with both of you. Thank you Aslı for giving me the freedom to go my own way, for your critical evaluation of my ideas, manuscripts, and presentations, and for your crucial role in creating and maintaining a gesture community in Nijmegen. Thank you Peter for making Nijmegen the best place in the world for neurobiological (and virtual) language research.

I sincerely thank the manuscript committee - Prof. Ton Dijkstra, Prof. Sotaro Kita, and Dr. Thom Gunter - for having read and evaluated this thesis. I greatly appreciate that you have been willing to spend time and effort into doing this. I would also like to thank Huib Kouwenhoven and Louise Schubotz for being there before and during the defense and, in general, for the fun times over the years.

During my time as a PhD-candidate, I have learned a lot from several senior researchers. Judith Holler generously shared her impressive knowledge of gesture research. Mingyuan Chu introduced me to experimental motion tracking. Tineke Snijders helped me set up an fMRI experiment and analyze the data. Roel Willems provided good advice and diverting side-projects along the way. Equally important have been Albert Russel, for expert help in the collection of motion tracking data and Paul Gaalman, who made fMRI data collection both fun and easy. I thank all of them for always being there exactly when needed.

Several people have contributed to creating stimulus materials and/or collecting data. Thank you Beyza, Doris, Flora, Huib, Hükü, José, Renske, and Reyhan for lending your voice,

face, and pointing hands, and Brenda, Charlotte, Laura, Livia, and Manu for help in data collection. Special thanks to Zeynep Azar for taking the early boat across the Bosporus to collect the Turkish data reported in this thesis, and to Prof. Aylin Küntay for kind local support.

I am grateful to the members of the Gesture and Sign Language lab and the Neurobiology of Language department for many nice meetings and constructive feedback throughout the years. Gwilym Lockwood and Ashley Lewis proofread parts of this thesis, and Linda Drijvers recommended aligning my text, which I greatly appreciate. Also the secretaries, administration, and technical groups of MPI and DCCN are gratefully acknowledged for their assistance, as well as Els, Rachel, and Dirkje for support from IMPRS.

Finally, I would like to thank my parents for interest and support, my brothers for the regular Brabant kind of entertainment, and the Deckers family for relaxing times in the resort. Most importantly, however, I thank Doris for her love.

Publications

Journal articles

Peeters, D., Snijders, T. M., Hagoort, P., & Özyürek, A. (under review). The neural integration of pointing gestures and speech in a visual context: An fMRI study.

Peeters, D., Chu, M., Holler, J., Hagoort, P., & Özyürek, A. (under review). Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech.

Peeters, D., Azar, Z., & Özyürek, A. (under review). The interplay between joint attention, physical proximity, and pointing gesture in demonstrative choice : Evidence from Dutch and Turkish.

Peeters, D., Hagoort, P., & Özyürek, A. (2015). Electrophysiological evidence for the role of shared space in online comprehension of spatial demonstratives. *Cognition*, 136, 64-84. doi:10.1016/j.cognition.2014.10.010.

Peeters, D., Runnqvist, E., Bertrand, D., & Grainger, J. (2014). Asymmetrical switch costs in bilingual language production induced by reading words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(1)*,* 284-292. doi: 10.1037/a0034060.

Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language, 68,* 315-332. doi:10.1016/j.jml.2012.12.003.

Dufau, S., Duñabeitia, J. A., Moret-Tatay, C., McGonigal, A., **Peeters, D.**, Alario, F.-X., Balota, D. A., Brysbaert, M., Carreiras, M., Ferrand, L., Ktori, M., Perea, M., Rastle, K., Sasburg, O., Yap, M. J., Ziegler, J. C., & Grainger, J. (2011). Smart phone, smart science: how the use of smartphones can revolutionize research in cognitive science. *PLoS ONE 6(9): e24974.* doi:10.1371/journal.pone.0024974.

Proceedings papers

Peeters, D., Snijders, T. M., Hagoort, P., & Özyürek, A. (2015). The role of left inferior frontal gyrus in the integration of pointing gesture and speech. In Proceedings of the 4th GESPIN - Gesture and Speech in Interaction - Conference. Nantes, France.

Peeters, D., Azar, Z., & Özyürek, A. (2014). The interplay between joint attention, physical proximity, and pointing gesture in demonstrative choice. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Meeting of the Cognitive Science Society* (pp. 1144-1149). Austin, TX : Cognitive Science Society.

Peeters, D., Chu, M., Holler, J., Özyürek, A., & Hagoort, P. (2013). Getting to the point: The influence of communicative intent on the kinematics of pointing gestures. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Meeting of the Cognitive Science Society* (pp. 1127-1132). Austin, TX : Cognitive Science Society.

Public Outreach

Peeters, D. & Dresler, M (2014). The scientific significance of sleep-talking. *Frontiers for Young Minds*, 2:9. doi: 10.3389/frym.2014.00009.

Peeters, D., Vanlangendonck, F., & Willems, R. M. (2012). Bestaat er een talenknobbel? Over taal in ons brein. In M. Jansen & M. Boogaard (Eds.), *Alles wat je altijd al had willen weten over taal. De taalcanon* (pp. 41-43). Amsterdam: Meulenhoff.

Curriculum vitae

David Peeters (Eindhoven, The Netherlands, 1987) graduated in Philosophy (BA), Communication- and Information Sciences (BA, MA), and Cognitive Neuroscience (M.Sc. cum laude). In 2010, he spent an academic year working on bilingual visual word recognition and language switching with Jonathan Grainger in Marseille, France. In 2011, he was awarded a 3-year doctoral IMPRS fellowship to carry out the work reported in this thesis. Currently he is a research staff member in the Neurobiology of Language department of the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands.

MPI series in psycholinguistics

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing. *Miranda van Turennout*
2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsagittal articulography. *Niels O. Schiller*
3. Lexical access in the production of ellipsis and pronouns. *Bernadette M. Schmitt*
4. The open-/closed-class distinction in spoken-word recognition. *Alette P. Haveman*
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach. *Kay Behnke*
6. Gesture and speech production. *Jan-Peter de Ruiter*
7. Comparative intonational phonology: English and German. *Esther Grabe*
8. Finiteness in adult and child German. *Ingeborg Lasser*
9. Language input for word discovery. *Joost van de Weijer*
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe. *James Essegbey*
11. Producing past and plural inflections. *Dirk J. Janssen*
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea. *Anna Margetts*
13. From speech to words. *Arie H. van der Lugt*
14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language. *Eva Schultze-Berndt*
15. Interpreting indefinites: An experimental study of children's language comprehension. *Irene Krämer*
16. Language-specific listening: The case of phonetic sequences. *Andrea C. Weber*
17. Moving eyes and naming objects. *Femke F. van der Meulen*
18. Analogy in morphology: The selection of linking elements in Dutch compounds. *Andrea Krott*
19. Morphology in speech comprehension. *Kerstin Mauth*
20. Morphological families in the mental lexicon. *Nivja H. de Jong*
21. Fixed expressions and the production of idioms. *Simone A. Sprenger*
22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria). *Birgit Hellwig*
23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies. *Fermín Moscoso del Prado Martín*
24. Contextual influences on spoken-word processing: An electro-physiological approach. *Daniëlle van den Brink*
25. Perceptual relevance of prevoicing in Dutch. *Petra M. van Alphen*
26. Syllables in speech production: Effects of syllable preparation and syllable frequency.

Joana Cholin

27. Producing complex spoken numerals for time and space. *Marjolein H. W. Meeuwissen*
28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction. *Rachèl J. J. K. Kemps*
29. At the same time...: The expression of simultaneity in learner varieties. *Barbara Schmiedtová*
30. A grammar of Jalonke argument structure. *Friederike Lüpke*
31. Agrammatic comprehension: An electrophysiological approach. *Marlies Wassenaar*
32. The structure and use of shape-based noun classes in Miraña (North West Amazon). *Frank Seifart*
33. Prosodically-conditioned detail in the recognition of spoken words. *Anne Pier Salverda*
34. Phonetic and lexical processing in a second language. *Mirjam Broersma*
35. Retrieving semantic and syntactic word properties. *Oliver Müller*
36. Lexically-guided perceptual learning in speech processing. *Frank Eisner*
37. Sensitivity to detailed acoustic information in word recognition. *Keren B. Shatzman*
38. The relationship between spoken word production and comprehension. *Rebecca Özdemir*
39. Disfluency: Interrupting speech and gesture. *Mandana Seyfeddinipur*
40. The acquisition of phonological structure: Distinguishing contrastive from noncontrastive variation. *Christiane Dietrich*
41. Cognitive cladistics and the relativity of spatial cognition. *Daniel B.M. Haun*
42. The acquisition of auditory categories. *Martijn Goudbeek*
43. Affix reduction in spoken Dutch. *Mark Pluymaekers*
44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence. *Valesca Kooijman*
45. Space and iconicity in German Sign Language (DGS). Pamela Perniss
46. On the production of morphologically complex words with special attention to effects of frequency. *Heidrun Bien*
47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture. *Amanda Brown*
48. The acquisition of verb compounding in Mandarin Chinese. *Jidong Chen*
49. Phoneme inventories and patterns of speech sound perception. *Anita Wagner*
50. Lexical processing of morphologically complex words: An informationtheoretical perspective. *Victor Kuperman*
51. A grammar of Savosavo, a Papuan language of the Solomon Islands. *Claudia Wegener*
52. Prosodic structure in speech production and perception. *Claudia Kuzla*
53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production

and comprehension. *Sarah Schimke*

54. Studies on intonation and information structure in child and adult German. *Laura de Ruiter*
55. Processing the fine temporal structure of spoken words. *Eva Reinisch*
56. Semantics and (ir)regular inflection in morphological processing. *Wieke Tabak*
57. Processing strongly reduced forms in casual speech. *Susanne Brouwer*
58. Ambiguous pronoun resolution in L1 and L2 German and Dutch. *Miriam Ellert*
59. Lexical interactions in non-native speech comprehension: Evidence from electro-encephalography, eye-tracking, and functional magnetic resonance imaging. *Ian FitzPatrick*
60. Processing casual speech in native and non-native language. *Annelie Tuinman*
61. Split intransitivity in Rotokas, a Papuan language of Bougainville. *Stuart Robinson*
62. Evidentiality and intersubjectivity in Yurakaré: An interactional account. *Sonja Gipper*
63. The influence of information structure on language comprehension: A neurocognitive perspective. *Lin Wang*
64. The meaning and use of ideophones in Siwu. *Mark Dingemanse*
65. The role of acoustic detail and context in the comprehension of reduced pronunciation variants. *Marco van de Ven*
66. Speech reduction in spontaneous French and Spanish. *Francisco Torreira*
67. The relevance of early word recognition: Insights from the infant brain. *Caroline Junge*
68. Adjusting to different speakers: Extrinsic normalization in vowel perception. *Matthias J. Sjerps*
69. Structuring language : contributions to the neurocognition of syntax. *Katrien R. Segaert*
70. Infants' appreciation of others' mental states in prelinguistic communication: a second person approach to mindreading. *Birgit Knudsen*
71. Gaze behavior in face-to-face interaction. *Federico Rossano*
72. Sign-spatiality in Kata Kolok: how a village sign language of Bali inscribes its signing space. *Connie de Vos*
73. Who is talking? Behavioural and neural evidence for norm-based coding in voice identity learning. *Attila Andics*
74. Lexical processing of foreign-accented speech: Rapid and flexible adaptation. *Marijt Witteman*
75. The use of deictic versus representational gestures in infancy. *Daniel Puccini*
76. Territories of knowledge in Japanese conversation. *Kaoru Hayano*
77. Family and neighbourhood relations in the mental lexicon: A cross-language perspective. *Kimberley Mulder*

78. Contributions of executive control to individual differences in word production. *Zeshu Shao*
79. Hearing speech and seeing speech: Perceptual adjustments in auditory-visual processing. *Patrick van der Zande*
80. High pitches and thick voices: The role of language in space-pitch associations. *Sarah Dolscheid*
81. Seeing what's next: Processing and anticipating language referring to objects. *Joost Rommers*
82. Mental representation and processing of reduced words in casual speech. *Iris Hanique*
83. The many ways listeners adapt to reductions in casual speech. *Katja Poellmann*
84. Contrasting opposite polarity in Germanic and Romance languages: Verum focus and affirmative particles in native speakers and advanced L2 learners. *Giuseppina Turco*
85. Morphological processing in younger and older people: Evidence for flexible dual-route access. *Jana Reifegerste*
86. Semantic and syntactic constraints on the production of subject-verb agreement. *Alma Veenstra*
87. The acquisition of morphophonological alternations across languages. *Helen Buckler*
88. The evolutionary dynamics of motion event encoding. *Annemarie Verkerk*
89. Rediscovering a forgotten language. *Jiyoun Choi*
90. The road to native listening: Language-general perception, language-specific input. *Sho Tsuji*
91. Infants' understanding of communication as participants and observers. *Gudmundur Bjarki Thorgrímsson*
92. Information structure in Avatime. *Saskia van Putten*
93. Switch Reference in Whitesands. *Jeremy Hammond*
94. Machine learning for gesture recognition from videos. *Binyam Gebrekidan Gebre*
95. Acquisition of spatial language by signing and speaking children: a comparison of Turkish sign language (TID) and Turkish. *Beyza Sumer*
96. An ear for pitch: on the effects of experience and aptitude in processing pitch in language and music. *Salomi Savvatia Asaridou*
97. lncrementality and Flexibility in Sentence Production. *Maartje van de Velde*
98. Social learning dynamics in chimpanzees: Reflections on (nonhuman) animal culture. *Edwin van Leeuwen*
99. The request system in Italian interaction. *Giovanni Rossi*
100. Timing turns in conversation: A temporal preparation account. *Lilla Magyari*
101. Assessing birth language memory in young adoptees. *Wencui Zhou*
102. A social and neurobiological approach to pointing in speech and gesture. *David Peeters*